



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2008

Histoire interne du français (Europe) : lexique et formation des mots

Glessgen, M D

Other titles: Interne Sprachgeschichte des Fränzösischen (Europa): Wortschatz und Wortbildung

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-11785>

Book Section

Published Version

Originally published at:

Glessgen, M D (2008). Histoire interne du français (Europe) : lexique et formation des mots. In: Glessgen, Martin-Dietrich; Ernst, G; Schmitt, C; Schweickard, W. Manuel d'histoire linguistique de la Romania. Berlin: de Gruyter, Art. 237.

Romanische Sprachgeschichte Histoire linguistique de la Romania

Ein internationales Handbuch zur Geschichte
der romanischen Sprachen
Manuel international d'histoire linguistique
de la Romania

Herausgegeben von / Edité par
Gerhard Ernst · Martin-Dietrich Gleßgen
Christian Schmitt · Wolfgang Schweickard

3. Teilband / Tome 3

237. Histoire interne du français (Europe): lexique et formation des mots
Interne Sprachgeschichte des Französischen (Europa):
Wortschatz und Wortbildung

Walter de Gruyter · Berlin · New York
2008

237. Histoire interne du français (Europe): lexique et formation des mots Interne Sprachgeschichte des Französischen (Europa): Wortschatz und Wortbildung

1. Difficultés de structuration du vocabulaire
2. Nombre et fréquence d'usage des mots français
3. Les origines du vocabulaire français
4. Les mécanismes de transformation lexicale dans l'histoire du français
5. L'interprétation onomasiologique de l'histoire du vocabulaire français
6. Aspects diasystématiques et évaluatifs
7. Perspectives
8. Bibliographie

1. Difficultés de structuration du vocabulaire

Le vocabulaire d'une langue répertorie l'intégralité du savoir partagé d'une communauté de locuteurs. Son évolution montre mieux que toute autre source les transformations mentales et comportementales d'une société, sa différenciation, sa créativité et ses prouesses intellectuelles (cf. infra 4.1.1.). En revanche, il est exceptionnellement difficile sinon impossible de structurer la nature d'un tel vocabulaire d'un point de vue fonctionnel ou typologique. Tous les mots lexicaux d'une même catégorie grammaticale – dans les langues indoeuropéennes les quatre catégories des noms, verbes, adjectifs et adjectifs – partagent essentiellement les mêmes fonctions grammaticales et ne peuvent pas être hiérarchisés les uns par rapport aux autres. Le lexique s'oppose donc radicalement à la phonétique / phonologie et à la grammaire, qui disposent d'un nombre ré-

duit de formes dont les fonctionnalités peuvent être décrites en termes de structure ou d'oppositions fonctionnelles. Seule la formation des mots introduit par la récurrence de ses mécanismes un élément systémique, ce qui la place à mi-chemin entre le lexique et la grammaire; étant donné la forte variabilité de l'inférence sémantique opérée par la dérivation et la composition, il est néanmoins plus cohérent de traiter la formation des mots parmi les mécanismes lexicologiques plutôt que comme élément de la grammaire: chaque dérivé connaît sa propre histoire sémantique et son propre ancrage diasystématique et syntagmatique, même s'il s'inscrit dans un type de formation fréquent et transparent dans le décodage.

La description globale du lexique d'une langue comme le français, suppose avant tout l'inventaire de dizaines de milliers de formes, chargées de polysémies et de liens valentiels et phraséologiques; l'historique d'un tel lexique demande l'observation de cette quantité de formes à travers mille ou deux mille ans d'histoire; l'identification de particularités typologiques enfin suppose la comparaison du français avec d'autres langues romanes ou non-romanes et des études analogues pour celle-ci (cf. les paramètres d'un traitement typologique du vocabulaire de Koch 2001, dans une optique essentiellement onomasiologique).

Il est vrai que les langues romanes possèdent cette particularité, mise en avant tout

au long de ce manuel, de permettre au moins en principe une description générale de leur vocabulaire dans son histoire, grâce aux sources disponibles, doublées de la documentation du lexique latin, excellemment répertorié par le *Thesaurus Linguae Latinae* (ThLL). Il s'ajoute dans le cas concret du français la qualité exceptionnelle des répertoires lexicographiques, avant tout avec le FEW qui embrasse la diachronie des langues galloromanes dans leur intégralité et dans une optique comparatiste (→ art. 32.2.); ces ouvrages fournissent au delà de l'existence des simples sources une base descriptive suffisante d'un point de vue épistémologique pour des interprétations lexicologiques globalisantes. Parallèlement à l'italien (→ art. 232), c'est un cas tout-à-fait à part parmi les langues du monde qui ouvre des perspectives uniques pour la recherche future; mais il reste malgré tout la difficulté de gérer des masses de données interprétatives considérables. Etant donné les énergies nécessaires pour de telles opérations et les caractéristiques propres à notre discipline (→ art. 1, 1.2.; 4.), il est possible qu'il s'agisse là d'une difficulté insurmontable, même à terme; dans tous les cas, jusqu'ici toute tentative d'affronter le vocabulaire français d'un point de vue fonctionnel, typologique ou globalisant est restée ponctuelle et limitée dans sa portée.

Sur la base de cette tradition d'études, nous présenterons par la suite une série de vues méthodologiques ou empiriques différentes dont chacune est très partielle mais dont l'ensemble donnera une idée certaine de la physionomie particulière du lexique français et de ses trajectoires évolutives. Il est particulièrement utile d'intégrer dans ces aperçus les paramètres de la fréquence d'usage, en distinguant les mots à fréquence forte, moyenne et faible, et des restrictions contextuelles, en regroupant les mots liés à certains genres textuels et portant éventuellement des marques diasystématiques. L'étude de certains champs sémantiques à une époque donnée ou en diachronie est plus traditionnelle en romanistique que la prise en considération de la fréquence et des contextes d'usage, même si, en dernière instance, les différents points de vue se chevauchent dans les faits. Nous devons accepter par ailleurs l'impossibilité épistémologique de présenter une vue d'ensemble pour les différentes époques de la langue comme pour la phonologie (→ art. 235) ou la grammaire (→ art. 236).

Notre présentation intègre une partie importante de l'argumentation et des données du chapitre sur le lexique dans Gleßgen (2007; sans renvois de détail). Par ailleurs, elle structure nouvellement les apports des nombreux autres articles du manuel à prendre en considération notamment pour les domaines des divers emprunts (y inclus les latinismes) et de la variation diasystématique (surtout diaphasique). Dans les trois domaines plus particuliers de la série des articles de lexicologie historique – la vue d'ensemble, le changement sémantique et la formation des mots –, les textes sur le français, sur l'italien (→ art. 232) et sur le roumain (→ art. 224a) ont été conçus parallèlement avec une volonté de complémentarité méthodologique. Les indications bibliographiques présentes dans les autres articles du manuel n'ont pas été reprises ici, sauf rares exceptions.

2. Nombre et fréquence d'usage des mots français

Le cadre de référence pour toute réflexion sur le lexique est créé par l'inventaire total des mots à un moment historique donné et par les conditions d'usage de chacun d'entre eux (fréquence selon les contextes, ancrage syntagmatique, marques diasystématiques). Il est possible de quantifier cet inventaire approximativement pour l'époque actuelle grâce à la lexicographie de référence du français:

- le PRob (2007) répertorie 52.645 entrées – la couverture du PRob (2004) annonce «60.000 mots» mais le décompte par catégories grammaticales dément ces chiffres; précisément: 39.384 noms, 12.046 adjectifs (dont 3.719 avec une valeur nominale qui n'ont pas été comptabilisés dans le décompte global), 1.673 verbes, 1.540 adverbes, 211 interjections, 264 mots grammaticaux, 228 locutions adjectivales / adverbiales. Le faible nombre de verbes (3%) et le grand nombre de substantifs (75%) soulignent la nature radicalement différente de ces deux catégories cognitivement et grammaticalement dominantes; dans ce sens, les verbes ont une dimension plus 'grammaticale' que les noms (importance des cadres valentiels, interactions multiples avec les différentes parties de la phrase, sémantisme plus abstrait et basé sur l'inférence, plus que sur des sèmes dénotatifs; cf. infra 4.3., n° 1);
- le TLF contiendrait d'après les différentes présentations entre 60.000 et 70.000 entrées;
- la plupart des entrées sont polysèmes, ce qui mène – d'après une extrapolation très grossière – à peut-être 250.000 sens lexicalisés (la couverture du PRob 2004 annonce «300.000 sens») dans ces deux dictionnaires qui rendent assez correctement compte de la langue actuelle, courante et soignée.

Il existe par ailleurs un vocabulaire diasystématiquement marqué comme par ex. les régionalismes de Suisse ou de France (ca. 1.000 entrées dans le DSR comme dans le DRF – mais auxquelles s'ajoute pour la seule Alsace ca. 350 nouvelles entrées dans le DRFA) ou le vocabulaire des jeunes et des banlieues (cf. infra 6.1.), des mots peu considérés par les dictionnaires comme, par ex., les onomatopées (ca. 300 entrées dans le *Dictionnaire des onomatopées*, Enckell / Rézeau 2003) ainsi que bon nombre de mots techniques dans les sciences actuelles (comme en linguistique le terme *oïlique*), sans parler d'un certain nombre de mots à très faible fréquence qui ne sont pas répertoriés dans ces dictionnaires. Mais il est peu probable que les mots lexicalisés du français actuel dépassent de beaucoup 100.000 unités formelles, sans compter toutefois les nomenclatures stricto sensu notamment des sciences exactes (cf. infra 6.3.) ni les noms propres et sans considérer les mots occasionnels dont le nombre est potentiellement illimité. L'inventaire lexicalisé et répertorié par la lexicographie est plus important pour l'anglais et même pour l'italien (→ art. 232) que pour le français; mais un ensemble d'env. 60.000 formes semble néanmoins correspondre au vocabulaire courant – et donc au savoir partagé – dans une société moderne alphabétisée.

Il n'est pas possible d'obtenir des chiffres aussi précis pour les époques anciennes puisque les genres textuels transmis opèrent toujours des restrictions d'usage. Cela ressort clairement des évaluations du vocabulaire des textes littéraires, connus par ailleurs pour une assez grande diversité lexicale en comparaison avec les genres textuels documentaires ou d'un savoir spécialisé. L'œuvre complète des auteurs les plus prolifiques atteint un maximum de 20.000 mots (pour le français, c'est le cas de La Fontaine ou Molière; l'œuvre de Rabelais comptabilise plus de 30.000 entrées mais en intégrant de nombreux termes techniques et des mots à diffusion réduite; cf. Brunet 2004, *Etendue du vocabulaire*: 33.328 lexèmes dont néanmoins 19.479 hapax); l'œuvre des auteurs classiques qui visent une langue stylisée et normalisée ne dépasse pas 4.000 mots (par ex. Corneille; Racine 3.263 mots; cf. Wolf ²1991, 125).

Tous les genres textuels réunis donneraient sans doute une idée approximative du vocabulaire des époques anciennes mais puisque la lexicographie historique a tou-

jours favorisé les textes littéraires – exception faite, notamment, du *Dictionnaire de Godefroy* (Gdf) –, nous ne disposons pas de données chiffrées sûres. En guise de comparaison, il est possible de prendre comme référence le vocabulaire du latin: le ThLL, dictionnaire exhaustif s'il en est, contiendra, une fois terminé, au maximum 50.000 entrées. Les langues romanes conservent par ailleurs un certain nombre de mots de toute évidence latins mais non attestés par les sources; ce vocabulaire utilisé exclusivement à l'oral augmenterait les chiffres d'env. 10% (cf. Stefenelli 1992, 22); en revanche, le ThLL contient bon nombre de mots rares, dont certains, sont sans doute des formations occasionnelles, et il couvre presque un millénaire d'histoire linguistique dont au moins cinq siècles qui disposent d'une documentation dense. Si le vocabulaire du français actuel est plus important que le vocabulaire du latin classique, cela n'est pas dû aux lacunes de la transmission mais au fait que la diversification du savoir ait entraîné une élaboration et créativité lexicale forte.

Par ailleurs, la fréquence d'usage de la plupart des mots est extrêmement faible: les 100 mots les plus fréquents couvrent plus de 60% des occurrences dans un corpus textuel équilibré, les 800 mots suivants presque 30% des occurrences; tous les autres mots de la langue se partagent les 10% restants: 8% pour les 5.800 mots suivants, 2% pour les ca. 65.000 mots (essentiellement noms et adjectifs) qui forment l'essentiel du corps des dictionnaires (cf. Wolf ²1991, 50; 182; Brunet 1981).

Parmi les mots à très haute fréquence se trouve notamment une bonne part des mots grammaticaux – en tout environ une centaine (dans les langues indoeuropéennes: pronoms, déterminants, prépositions et conjonctions) – dont l'étude appartient (contrairement à celle de la formation des mots) pleinement à la grammaire. Les mots lexicaux à haute fréquence sont pour la plupart très polysèmes et disposent, dans le cas des verbes, de cadres valentiels diversifiés. Les mots à très faible fréquence en revanche sont souvent porteurs de sens très spécifiques, liés aux concepts d'une culture élaborée, ce qui rend illusoire toute tentative de limiter l'interprétation du vocabulaire d'une langue aux seuls mots fréquents. Les caractéristiques et les comportements syntaxiques des mots changent donc considérablement selon la tranche de fréquence à laquelle ils appartiennent.

3. Les origines du vocabulaire français

3.1. Aperçu quantitatif

La structuration étymologique du vocabulaire français par langues d'origine est le seul domaine lexicologique qui permet des quantifications relativement solides. Celles-ci restent toutefois partielles puisqu'elles portent exclusivement sur le nombre de formes lexicales et, dans une certaine mesure, sur leur fréquence d'usage dans des textes choisis; elles n'impliquent pas le nombre de sens lexicaux ni leur cadre valentiel, les collocations et phraséologismes ni encore les sous-ensembles diasystématiques; mais il s'agit néanmoins de quantifications globalisantes. L'établissement de la filiation des mots par l'étymologie permet par ailleurs de mieux structurer les changements lexicaux et même le vocabulaire en synchronie.

L'essentiel du vocabulaire français provient naturellement du latin mais avec cette particularité que certains mots sont de type héréditaire, d'autres sont le résultat d'interférences avec le latin comme langue écrite dominante ou de référence. Puisque le latin constituait dans la Romania jusqu'au XVI^e s. une variété de prestige à l'intérieur d'un même continuum communicatif, il est présent dans les formes grapho-phonétiques des mots, dans leur sémantisme et dans les morphèmes dérivationnels (→ art. 135; 136; 259, 3.; Wolf 1991, 156–154); les multiples traductions ont notamment renforcé l'impact du modèle linguistique latin (→ art. 121, 5.1.). Les emprunts 'isoglottes' au latin continuent jusqu'à l'époque récente qui connaît de nombreux internationalismes dans les langages techniques et scientifiques (→ art. 70, 6.; 258, 2.2.). Les emprunts à d'autres langues de contact sont nettement plus réduits, autant en termes de nombre de mots qu'en termes de fréquence d'usage.

Des quantifications précises et solides ne sont jamais très faciles à obtenir. Les premières listes fouillées de mots à haute fréquence furent établies par Gougenheim et al. (1956; 1964), suivies du dictionnaire de Juilland / Brodin / Davidovitch (1970) qui reste méthodologiquement discutable (cf. la critique minutieuse d'Ernst 1979). Les recherches récentes en linguistique historique de corpus ouvrent d'importantes perspectives dans ce domaine (cf. Pusch / Kabatek / Raible 2005) mais avant l'élaboration de nouveaux travaux de référence, nous devons nous en tenir aux évaluations plus ancien-

nes; voici les relations numériques sur les éléments de formation du vocabulaire français établies par Wolf (1991, 50 s.; 182 s.) sur la base du dictionnaire de Juilland / Brodin / Davidovitch (1970):

- les 100 mots les plus fréquents sont tous des mots héréditaires dont seulement huit dérivés (56 mots grammaticaux, 19 adverbes, 16 verbes, 5 adjectifs et seulement 4 substantifs);
- les 2.400 mots suivants restent essentiellement d'origine latine (héréditaire ou savante); seuls 44 mots représentent des emprunts 'alloglottes' postérieurs à l'an mil (= 1,7%), à l'italien (27 mots), à l'anglais (11 mots) et à des langues de contact mineures (en tout 6 [ou 7] mots);
- les mots considérés comme 'usuels' (sans quantification) se diversifient:
 - > les emprunts alloglottes montent à 5 ou 6%, ce qui correspond assez précisément à leur fréquence d'usage dans un texte journalistique actuel;
 - > les formes non dérivées à évolution phonétique héréditaire représentent, elles aussi, seulement 5% (par ex. *père* < *PATRE*-, *mère* < *MA-TRE*-);
 - > les emprunts isoglottes au latin, non dérivés, s'élèvent à 25%;
 - > les dérivés sur la base de mots latins héréditaires ou savants correspondent à 65% des mots actuels (les bases héréditaires sont largement plus fréquentes).

Sur l'ensemble du vocabulaire actuel répertorié dans le PROB (2004) et dans le TLF, le pourcentage d'emprunts alloglottes double facilement (cf. infra 3.4.: notre décompte relève plus de 6.300 emprunts), ce qui souligne une nouvelle fois les difficultés inhérentes aux données chiffrées. Mais il est toutefois certain que le nombre d'emprunts (alloglottes et isoglottes) ainsi que celui des dérivés augmente à travers l'histoire du français. Cette augmentation se produit probablement de manière régulière, même si les différents mécanismes d'innovation changent en importance selon les époques (cf. par ex. pour le nombre des latinismes au MA → art. 136, 3.6.).

3.2. Les mots d'origine latine

Les mots d'origine latine fournissent l'immense majorité du lexique français actuel. Il est toutefois utile de distinguer les mots héréditaires non dérivés – sans ou avec changement sémantique –, les mots héréditaires dérivés et, en dernier lieu, les emprunts isoglottes au latin, ce qui est plus facile en français que dans d'autres langues romanes grâce aux changements phonétiques très développés. Ces distinctions sont opérées de manière

re systématique dans le FEW (opposition entre les mots héréditaires regroupés sous «I.» et les mots (semi-)savants regroupés sous «II.», distinction de mots simples et de mots construits). Elles permettent dans un deuxième temps de formuler des hypothèses pour une interprétation de type cognitif. Il est probable que les mots héréditaires non dérivés – au nombre de 100 termes parmi les 927 mots lexicaux les plus fréquents (Stefenelli 1981, 10, n. 4; 254) – représentent des concepts centraux dans le réseau sémantique neuronal, pour des raisons anthropologiques ou culturelles. Ils correspondent souvent à des concepts du monde de références quotidiennes; cf. le relevé des substantifs d'après Stefenelli (ib., 255–301):

identité: *femme, homme, père, mère, parents, fils, fille, sœur, mari, oncle, ami, voisin, Dieu; ouvrier, maître*

condition humaine: *vie, mort, âge, santé, douleur, peine, faim, chaleur*

sens: *langue, voix, œil, nez, main, pied*

univers: *terre, mer, air, île, neige*

temps: *temps, an, mois, heure, été, soir, nuit*

aliments: *eau, pain, vin, lait, œuf*

objets de l'habitat: *mur, fenêtre, lit, verre, sac, fil*

monde rural et commercial: *champ, bête, chien,*

fleur, laine, marché, prix

loisirs: *jeu, bain, chant*

description du monde: *nom, nombre, ordre, mesure,*

ligne, point, part, lieu, fin, cours, couleur

évaluation: *fait.*

S'ajoute des verbes fondamentaux comme *avoir, dire, voir, venir, mettre, devoir, tenir, laisser, aimer, appeler, entendre*, des adjectifs (*grand, bon, meilleur, vieux, jeune* etc.) et adverbes (*mieux, moins, mal* etc.). Les lexèmes cités ont tous gardé au moins un de leurs sens latins d'origine, ce qui confirme qu'il s'agit de concepts particulièrement stables à travers les époques (cf. aussi Wolf 21991, 52).

D'autres termes, toujours du monde des références quotidiennes, sont en revanche le produit d'un changement sémantique, parfois notable comme la métaphore connue *tête* < TESTA "pot; tesson", mais souvent peu spectaculaire comme la métonymie *ville* < VILLA "maison ou propriété rurale". Certains de ces concepts sont sémantiquement très actifs en tant que centres d'expansion métaphorique ou métonymique (all. *Bildspender*); différentes études romanistes ont prouvé, notamment, la grande productivité du concept de la "tête" (cf. infra 4.1.2.). Les formes caractérisées par des innovations

sémantiques sont plus fréquentes que les formes sémantiquement stables mais elles se chevauchent souvent avec le troisième groupe de mots héréditaires qui ont fait l'objet d'une dérivation.

Les dérivés interviennent à partir de l'époque du latin tardif (comme *oreille* < AURICULA ← AURIS, *pays* < *PAGENSE, *contrée* < (REGIO) CONTRATA "(région) en face, devant les yeux" ou *viande* < VIVENDA "ce qui est nécessaire pour vivre") mais ils restent présents à l'époque médiévale et moderne (*marier* → *mariage* [XII^e s.], *vite* → *vitesse* [XII^e s.]; cf. les composés *jeune fille* [XV^e s.] ou *après-midi* [XVI^e s.]; cf. encore la synthèse de Stefenelli 1981, 155–301, avec de nombreux exemples d'innovations sémantiques et dérivationnelles et → art. 50, 4.4.). L'importance de la dérivation et de la composition augmente largement pour les mots à fréquence moyenne et faible.

Les différentes catégories des mots héréditaires s'opposent aux latinismes, surtout présents parmi les mots à fréquence moyenne ou faible et parmi les concepts liés à la culture élaborée; cependant parmi les 927 mots à haute fréquence étudiés par Stefenelli (1981), env. 150 termes ont été empruntés au latin ou relatinisés dans leur forme ou / et dans leur sens à travers les siècles; cf. par ex. les noms:

épouse (XI^e, < SPONSA), *famille* (XIII^e, < FAMILIA) *service* (XI^e, < SERVITIUM)

étude (XII^e, < STUDIUM), *histoire* (XII^e, < HISTORIA)

genre (XII^e, < GENUS), *espèce* (XIII^e, < SPECIES),

différence (XII^e, < DIFFERENTIA)

région (XII^e, < REGIO), *état* (XII^e, < STATUS)

opinion (XII^e, < OPINIO), *moment* (XII^e, < MOMENTUM)

musique (XII^e, < MUSICA); cf. pour d'autres ex. → art. 136; 258, 2.2.1.

Parmi les mots à fréquence moyenne et faible, les latinismes occupent une place toujours croissante. Le grand nombre de latinismes explique l'extrême complexité de leur histoire dans les langues romanes. À côté des emprunts intégraux, il existe de nombreux cas de relatinisation formelle ou sémantique de mots à évolution héréditaire, de doublons étymologiques (par ex. *frêle* – *fragile*, < FRAGILIS) ou de familles de mots constituées à la fois de formes héréditaires et savantes (par ex. *eau* – *aqueux*, → art. 135; 136; 232, 2.).

Un sous-ensemble particulièrement épineux parmi les mots savants est constitué par les hellénismes du latin: le grec était la deuxième langue de l'empire romain bilin-

gue. La recherche lexicologique classe ces lexèmes souvent à tort comme 'grécismes', ce qui est vrai par rapport au latin et comme *etimologia remota*; mais par rapport au français, il s'agit simplement de latinismes, même s'ils peuvent porter les marques reconnaissables de leur origine lointaine (par ex. les digraphes *ph* ou *th*). Jusqu'à la Renaissance, pratiquement tous les 'grécismes' passent par le latin. Ils alimentent également par là une bonne partie de la terminologie scientifique et technique depuis le XIX^e s. (→ art. 138, 3.); ici, les éléments de formation ('confixes') jouent un rôle tout aussi important que les formes intégrales (→ art. 232, 4.3.). Les hellénismes sont presque absents des mots à haute fréquence (cf. PARABOLA [fr. *parole*] et COLAPHUS [fr. *coup*], transmis par voie héréditaire, ainsi qu'IDEA, seul mot savant dans cette catégorie) mais augmentent en importance dans les autres tranches de fréquence. Depuis le XVI^e s., les emprunts directs au grec et les emprunts au latin coexistent intimement; cf. quelques exemples pour la lettre A-, tous transmis par le latin:

alphabet (ca. 1140 < ALPHABETUM < gr. *alpha* + *bêta*)

absinthe (1190 [*absince*] < ABSINTHIUM < gr. *apsinthion*)

aimant (XII^e s. [*aiemant*] < *ADIMAS < ADAMAS < gr. *adamas*)

anatomie (1370, < ANATOMIA < gr. *anatomia*)

acacia (XIV^e s. [*acace*] < ACACIA < gr. *akakia*)

analogie (1528, < ANALOGIA < gr. *analogia*)

acore (XVI^e s. < ACORUM < gr. *akoros*)

anecdote (1685, < ANECDOTA pl. < gr. *anekdota*) etc. etc. (→ art. 136, 4.6.; 138, 3.1.; 258, 2.2.2.).

Le choix entre un mot héréditaire ou savant dans l'histoire du français ne répond pas toujours à une raison sémantique probante; l'introduction d'un terme latin est normalement un effet d'élaboration linguistique, parfois aussi un simple reflet du grand prestige de cette langue savante et ecclésiastique (cf. Stefenelli 1992, 205ss.). L'interprétation du rôle de ces différentes catégories dans l'histoire du français est toujours délicate et il manque des études systématiques dans une optique cognitive. Néanmoins, le français comme d'autres langues romanes permet de telles interrogations d'un intérêt réel pour la linguistique générale. La grande présence de latinismes confère à ces langues une particularité sinon typologique mais physiologique.

Un dernier point, parmi les raisonnements très généraux, mérite l'attention: le choix

opéré parmi les mots latins lors du passage vers les langues romanes. Là encore, la fréquence joue un rôle primordial (cf. ib., 31s.; 35):

- parmi les 1.200 lexèmes les plus fréquents en latin écrit, la moitié survit dans plusieurs langues romanes (notamment les mots non dérivés concernant des références quotidiennes, cités supra pour le français), 100 autres lexèmes dans une seule langue romane;
- parmi les mots usuels du latin, 50% survivent dans au moins une langue romane;
- les mots rares du latin écrit disparaissent presque tous de l'oral, ce qui fait que la base latine héréditaire des langues romanes ne représente que 10% des mots attestés par les sources (sachant que celles-ci attestent bien 80% des mots romans d'origine latine, cf. supra);
- enfin, grâce aux emprunts isoglottes au latin, plus d'un tiers du vocabulaire latin attesté finit par vivre aujourd'hui dans au moins une langue romane.

La place du français dans ce scénario n'est pas particulière par rapport à la plupart des autres langues romanes, même si dans le détail les pourcentages varieront (cf. aussi pour la régionalisation du vocabulaire du latin tardif → art. 49, 3.3.3.); mais le français comme chacune des langues romanes retient bien moins de 10% du vocabulaire latin parmi les mots héréditaires (cf. Stefenelli 1992, 94ss., et n. 164 qui énumère une trentaine de lexèmes latins caractéristiques pour le français à l'intérieur de la Romania comme CARMEN, VOCARE, INSTAURARE; cf. aussi Stefenelli 1996, 370ss.). Les langues romanes opèrent donc un choix lexical radical à l'époque de leur genèse, tout en élargissant ce choix très rapidement par la dérivation et le changement sémantique. Les comptes sont faussés par l'impact du latin écrit qui augmente considérablement la stabilité du vocabulaire dans le temps. Mais malgré cet effet stabilisateur, le cas du français montre combien le vocabulaire peut se transformer en profondeur à travers seulement deux mille ans.

3.3. Les emprunts alloglottes anciens (avant l'an mil)

Le vocabulaire héréditaire du français inclut en dernière instance tous les emprunts que le latin a opéré jusqu'au VII^e s. sur le territoire d'oïl. Un locuteur analphabète du X^e s. ne pouvait pas différencier un mot d'origine celtique ou grecque d'un mot d'origine latine; il est même douteux que les emprunts

germaniques des VI^e / VIII^e s. aient été encore reconnaissables en tant que telles vers l'an mil.

Les lexèmes du latin antique provenant de langues alloglottes ne sont pas très fréquents (cf. pour les emprunts en onomastique, plus nombreux → art. 239, 2.2.1.). Le plus grand nombre d'emprunts concerne des grécismes, notamment dans le vocabulaire savant (cf. supra 3.2.). En revanche, les grécismes par contact de population sont très rares dans la Galloromania et encore plus en français.

Les emprunts celtiques sont les plus fréquents parmi des interférences avec les autres langues de contact du latin. Ils participent même à la différenciation régionale du latin et bon nombre de mots celtiques sont essentiellement galloromans. En tout, 330 étymons de mots galloromans ont été considérés comme gaulois, souvent sans doute à tort (→ art. 55, 1.4.; 70, 3.). Les mots français d'origine gauloise concernent pour la plupart des concepts de la vie rurale et de la description du territoire comme *arpent*, *boue*, *bruyère*, *chemin*, *lie*, *mouton*, *ruche* (→ art. 55, 1.4.).

Les autres langues de 'substrat' restent nettement plus difficiles à cerner (en tout env. 80 entrées dans le FEW, souvent discutables, cf. ib.).

Les emprunts aux langues germaniques sont bien plus nombreux que tous les autres emprunts anciens par contact de population; ils se placent pour la plupart dans les deux siècles après les invasions germaniques, aux VI^e et VII^e s., même si certains sont plus anciens et même si le territoire d'oïl devait connaître encore plus tard une noblesse franque bilingue (→ art. 55, 2.1.). Ces interférences interviennent à l'époque précise de l'émergence des langues romanes, ce qui explique pourquoi a pu germer l'idée que les langues germaniques auraient déterminées comme 'superstrats' la genèse de la Romania, même si cette hypothèse s'est avérée erronée (→ art. 70, 4.). L'impact du francique sur le français est toutefois important, surtout en onomastique (→ art. 239) et dans le lexique: le FEW dédie trois volumes aux germanismes contre actuellement seize volumes pour les mots d'origine latine (y inclus les hellénismes et les quelques celtismes). Ces relations sont à considérer avec prudence, puisque les trois volumes de mots d'origine inconnue sont difficilement quantifiables et puisque les germanismes ont connu un traitement particulièrement intense dans le

FEW, Walther von Wartburg étant l'auteur principal de la théorie des superstrats. Il n'en reste pas moins que le français ancien et les dialectes d'oïl se caractérisent par un pourcentage de germanismes bien supérieur à celui de la langue standard moderne; la relatinisation permanente du français a affaibli par la suite l'élément germanique.

Les emprunts germaniques (725 étymons dans le FEW dont 500 franciques, → art. 55, 2.4.) se concentrent dans des champs sémantiques tels que la vie militaire et princière mais aussi rurale, comme par ex.:

germ. **baro* "homme libre servant à l'armée" > fr. *baron* "homme brave, valeureux" puis "grand seigneur du royaume" (X^e s.)

afreq. **sparwari* "épervier" > *épervier* (1080 [*esprevier*])

afreq. **werra* "combat" > fr. *guerre* (dp. XI^e s.)

afreq. **siniskalk* "maître d'hôtel" > fr. *sénéchal* "officier féodal, chef de la justice" (1119 [*senechal*])

a.b.freq. **waidanjan* "paître, mener les bêtes aux champs" > fr. *gagner* (XII^e s.).

Malgré ces restrictions sémantiques, il est légitime de rapprocher les mots d'origine germanique des mots héréditaires d'origine latine (→ art. 70, 4.). Comme ces derniers, ils ont connu des évolutions dérivationnelles et sémantiques ultérieures bien plus marquées que les emprunts alloglottes postérieurs.

3.4. Les emprunts alloglottes après l'an mil

La très grande majorité des emprunts postérieurs à l'an mil est parvenue au français à partir d'une autre langue romane ou encore à travers le latin médiéval ou moderne. Les emprunts arabes passent au MA à travers l'espagnol, l'italien et le latin, à son tour en interaction avec les deux langues romanes méridionales; les termes amérindiens à l'époque moderne viennent surtout par l'espagnol. Les arabismes et les amérindianismes restent néanmoins souvent reconnaissables en tant qu'emprunts d'origine exotique par leur forme et leur sémantisme, au moins à l'époque de l'emprunt. Les premiers se concentrent dans les domaines des produits exotiques et des sciences médiévales, comme par ex.:

ar. *az-zaʿfarān* > lat.méd. *safranum* > fr. *safran* (dp. XII^e s.)

ar. (*as-*)*samt* (*ar-raʿs*) > lat.méd. *zemt* → lat. méd.it. (?) *zenit* (la faute de lecture *m* → *ni* semble avoir eu lieu dans des textes latins d'Italie) > esp. *cenit* > fr. *cenith* (XIV^e s., Oresme), *zenith* (pour d'autres ex. → 143, 1.2.; cf. aussi 155, 1.2.).

Les derniers concernent essentiellement des concepts caractéristiques de l'Amérique ibérique à l'époque coloniale (→ art. 70, 5.), par ex.:

caraïbe *kanawa* > esp. *canoa* > fr. *canoë* (dp. 1519)
Haïti *hamaca* > esp. ~ > fr. *amacca* (1533), *hamac*
náhuatl *tomatl* > esp. *tomate* > fr. ~ (dp. 1598).

Les emprunts directs à l'espagnol restent par ailleurs peu fréquents en français (cf. au XVIII^e s. esp. *siesta* > fr. *sieste*). Très nombreux sont en revanche les italianismes de la Renaissance qui ont un impact réel dans les domaines sémantiques de la culture élaborée comme les arts, l'architecture et la musique ou la cuisine, de même que dans le vocabulaire financier et militaire (→ art. 70, 5.; 155, 1.3.; 258, 2.2.3.); cf. par ex.:

it. (*dipingere a*) *fresco* "(peindre sur un enduit frais)" → "œuvre murale peinte en utilisant des couleurs délayées à l'eau sur un enduit de mortier frais" > fr. *fresque* (dp. 1669)

it. *adagio* "à son aise, doucement" → "(musique) indication de mouvement lent" > fr. *adagio* (dp. 1726).

La deuxième langue de contact romane, en termes de nombre d'emprunts directs, est l'occitan qui se trouve depuis la fin du MA dans le même espace communicatif que le français (→ art. 70, 5.). Les évaluations chiffrées sont très contradictoires: Karl Gebhardt avance – en 1974 – le chiffre de 1.600 lexèmes, le TLFi indique un emprunt à l'occitan dans 276 entrées (d'après une évaluation de Patrick Sauzet, cf. Chambon / Carles 2007), le PRob (2007) 517; Gebhardt se base sur le FEW et prend en considération l'intégralité du diasystème français avec des formes anciennes et dialectales, le TLFi en revanche ne répertorie que les mots d'usage contemporains; le plus grand nombre d'entrées dans le PRob s'explique pour l'essentiel par un meilleur balisage (bon nombre des termes en question sont étiquetés dans le TLFi comme occitanismes mais n'apparaissent pas dans la liste des 'emprunts au provençal'), parfois par une nomenclature mieux choisie, parfois aussi par des décisions étymologiques divergentes (toujours plus solides dans le TLFi).

Par ailleurs, la conceptualisation des emprunts manque de cohérence dans tous les répertoires: les 'occitanismes' du français sont pour la plupart des emprunts opérés par une variété régionale du français depuis le XVI^e s. et diffusés plus tard dans la langue générale; il s'agit donc d'emprunts 'internes'

(cf. Chambon / Carles 2007) ou variationnels (cf. infra 6.2.). D'un point de vue sémantique, ces mots portent fréquemment sur des concepts liés d'une manière ou d'une autre au Sud de la France, la vie rurale et urbaine, l'environnement, la culture régionale ou la géographie physique (→ art. 155, 4.); par ex.:

occ. *abelha* > fr. *abeille* (XIII^e s.)

occ. *bastida* > fr. *bastide* (1355; le mot occ. provient à son tour du mot frq. **bastjan*)

occ. *banqueta* > fr. *banquette* (1417 [dans le sens "selle"])

occ. *rodar* > fr. *rôder* (1418 [sous la forme *rodder*])

occ. *garriga* > fr. *garrigue* (1544)

occ. *aioli* > fr. *aïoli* (1744).

L'exemple de l'occitan montre combien les quantifications d'emprunts sur la base du TLFi et du PRob (2007) sont sujettes à caution, en parfait contraste avec leur grande facilité d'accès. Pour les différentes langues alloglottes, une recherche ciblée fournit très rapidement les chiffres suivants:

anglais:	868 – 2.793 (= TLFi – PRob 2007)
italien:	668 – 1.199
occitan:	246 (sous les entrées "provençal", "occitan", "limousin" etc.) – 517
espagnol:	233 – 451
arabe:	217 – 416
allemand:	208 – 541
néerlandais:	125 – 245
portugais:	49 – 116
russe:	46 – 96 (pour d'autres quantifications → art. 70, 5.).

Pour l'anglais, les indications du PRob sont d'autant plus élevées qu'il s'agit de mots souvent récents voire postérieurs à l'établissement de la nomenclature du TLF; mais les différences ne s'arrêtent pas aux questions du balisage et de l'époque d'emprunt: toute vérification de détail montre que l'identification et la détermination lexicographiques des emprunts en français ne sont pas bien établies. Cet état de fait pourrait être motivé par la relative faiblesse du FEW dans ce domaine: les volumes 18 à 20, consacrés aux emprunts alloglottes après l'an mil, sont très succincts et ont connu depuis des compléments solides (Arveiller 1999 pour les mots orientaux; Buchi 1996, 165–257, pour les slavismes, → art. 141, 3., cf. aussi Baldinger 1988/2002 pour les mots d'origine incon nue). Par ailleurs, l'identification précise des emprunts du FEW demande souvent un certain effort; dans les vol. XV à XVII, par ex., les germanismes anciens apparaissent peu

distincts à côté des emprunts de la langue allemande standard, des emprunts par contact frontalier en Lorraine depuis le MA et des emprunts par contact linguistique récent en Alsace (→ art. 148, 2.4.; 4.2.; 164, 4.2.).

Quant aux autres langues mentionnées, le néerlandais intervient surtout au MA (dp. le XII^e s.) et à l'époque de l'Ancien Régime, particulièrement dans le vocabulaire du commerce et de la navigation (→ art. 148, 4.1.); par ailleurs, ce sont surtout les anglicismes et les arabismes récents qui méritent l'attention.

L'anglais dépasse de loin toutes les autres langues de contact depuis la deuxième moitié du XX^e s. (→ art. 150, 2.; 258, 2.2.3.; cf. les exemples infra 3.5.). Etant donné le grand nombre de mots anglais d'origine française (→ art. 65b, 2.4.; 161, 2.), ces emprunts représentent souvent des revenants (par ex. fr. (*auto-*)*car* ← angl. *car* ← norm. *car* < CARRUS ou fr. *sport* ← angl. ~ ← afr. *de(s)port*).

L'arabe intervient comme langue de contact directe essentiellement depuis la colonisation française du Maghreb (→ art. 143, 2.) mais son action s'est intensifiée dans le langage récent des banlieues et des jeunes, élaboré dans un genre textuel comme le rap (→ art. 206, 5.1.); cf. pour les emprunts d'époque coloniale:

ar. *tabib* > fr. *toubib* "médecin" (1863 [*tohib*]).
ar. *fi-s-sā'a* "à l'instant, vite" > fr. *fissa* adv. "vite" (1909)
ar. *b(i)lad* > fr. *bled* "petit village isolé, sans ressources" (1934 ["terrain, pays", fin XIX^e s.]).

Si les grands mouvements d'emprunts du français sont facilement identifiables, la question du cheminement précis des emprunts demande toujours une très grande attention dans le détail. Souvent, il peut même être difficile d'identifier avec sécurité la langue-source pour un emprunt comme dans le cas des arabismes médiévaux du français où la décision de l'*etimologia proxima* entre le latin médiéval, l'italien, l'espagnol et éventuellement le catalan n'est pas toujours possible. Dans le même ordre d'idées, le français a véhiculé lui-même de nombreux anglicismes vers l'espagnol ou le roumain modernes (cf. pour les nombreux emprunts d'autres langues au français → art. 152, 2.; 153c, 3.; 154; 156, 3.; 158; 224a, 10.; 253, 3.2.2.; 258, 2.2.3.; pour la situation particulière des emprunts vers le français en dehors de l'Europe → art. 238, 2.; 76, 2.4.; 77–81; 151, 1.).

3.5. Typologie des emprunts

Tous les mouvements d'emprunt provoquent des interférences relativement fortes à un moment historique donné dont les résultats s'affaiblissent par la suite; une fois terminée la phase d'accueil, les éléments étrangers sont en partie rejetés, en partie intégrés pour de bon. L'intégration suit une logique par étapes successives (illustrées ici par l'exemple des anglicismes, récents mais pour la même raison bien saisissables; cf. pour la typologie Schweickard 1998; → art. 150, 1.6.; 2.3.; pour les datations Höfler 1982):

- pour la plupart, les emprunts sont intégrés dans le système de prononciation de la langue-cible (adaptation phonétique): par ex. *jungle* [ʒœgl] < angl. *jungle* [dʒɒŋgl] (dp. 1796, FEW 18, 76a); les formes récentes peuvent connaître néanmoins des variations; par ex. pour *pipeline* (2^e moitié XIX^e s.) coexistent la prononciation francisée [pi'plin] et la prononciation plus proche de l'origine [pa'plajn];
- l'adaptation graphique ou aux règles de relations grapho-phonétiques est en principe parallèle à l'adaptation phonétique (par ex. *cornère* [Êtiemble, cf. TLFi] vs. *corner*); le jeu des deux formes d'adaptation peut mener à des transformations importantes comme dans *boulingrin* [bulɛ̃'grɛ] "emplacement gazonné pour le jeu de boules, pelouse rectangulaire, entourée de petits talus" < angl. *bowling-green* (1663, FEW 18, 34b);
- l'adaptation des morphèmes grammaticaux est inévitable pour les verbes: fr. *boycotter* (dp. 1880) vs. angl. *to boycott*; de même, l'adaptation morphologique flexionnelle intervient nécessairement pour le genre des noms (*la star* [dp. 1919] d'apr. *la vedette*); elle est plus variée pour les marques de nombre (*les boxes* vs. *les box*, dp. fin XVIII^e s.);
- l'adaptation morphologique dérivationnelle représente une intégration plus avancée, cf. *agnostique* (dp. 1884) ← angl. *agnostic* (adaptation au suffixe habituel *-(i)que*), *indésirable* (dp. 1801) ← angl. *undesirable* (adaptation au préfixe *in-*), *possiblement* (mot médiéval sorti d'usage et repris suite à l'influence anglaise) ← angl. *possibly*;
- la formation de nouveaux dérivés ou composés sur un emprunt n'est pas fréquente dans un premier temps et dénote une intégration ultérieure: fr. *hockey* (dp. 1876) → *hockeyeur* (dp. 1910);
- les emprunts connaissent souvent de fortes restrictions sémantiques: ils ne conservent généralement qu'une partie des sens de la forme d'origine (par ex. fr. *gay* "homosexuel" [dp. 1952] < angl. *gay* fortement polysémique) et ne développent des sens nouveaux qu'avec le temps;
- l'intégration des emprunts dans de nouveaux phraséologismes dénote comme la dérivation ou le changement sémantique un degré supérieur d'intégration (par ex. *être publié sur le web* [écrit aussi *ouèbe*]).

Les emprunts intégraux s'accompagnent toujours de nombreux emprunts sémantiques ou calques. Les emprunts de sens (comme fr. *arrière* "joueur dans la défense" d'apr. angl. *back*) peuvent être catalysés par l'intervention d'une proximité formelle (par ex. fr. *approche* "manière d'aborder un sujet" [dp. 1962?] d'apr. angl. *approach*, fr. *crucial* "décisif" [dp. 1911] d'apr. angl. *crucial*, fr. *réaliser* "se rendre compte" [dp. 1858] d'apr. l'angl. *to realize*). Les emprunts de structure sont moins fréquents (par ex. l'internationalisme *gratte-ciel* [dp. 1911] d'apr. *sky-scraper* ou, plus récemment, fr. (*faire une recherche*) *en ligne* [1989] d'apr. *on line*). Les emprunts sémantiques ou de structure sont nettement plus difficiles à détecter que les emprunts formels mais ils se trouvent, à l'époque de l'interférence, dans les mêmes situations communicatives et genres textuels que les emprunts de mots. Il est vrai que par la difficulté de leur perception – qui affecte même les locuteurs contemporains lors de l'emprunt – ces formes ne génèrent pas les mêmes jugements de valeur que les mots étrangers empruntés, considérés par certains comme 'beaux' et 'biens', par d'autres comme 'mauvais' et 'laid's' (cf. pour l'exemple bien étudié et très caractéristique des anglicismes: Schweickard 2005; cf. infra 6.4.).

Plus rarement, des emprunts impliquent des transformations dans les mécanismes préférentiels de formation de mots (comme les suffixes romans qui correspondent aux formes angl. *-ize* / *-ization* / *-ism* / *-ist*, → art. 70,7.; cf. aussi l'argumentation fine et complexe → art. 259) voire même dans les choix syntagmatiques ou syntaxiques.

L'action conjointe des différents types d'emprunts peut mener à la restructuration d'un champ sémantique défini. Malgré sa part restreinte dans le lexique, l'emprunt participe donc pleinement au processus d'élaboration linguistique, si caractéristique pour les langues de culture européennes du deuxième millénaire.

4. Les mécanismes de transformation lexicale dans l'histoire du français

4.1. Le changement sémantique

4.1.1. La théorie du changement sémantique

L'inventaire lexicologique hérité et emprunté connaît des transformations très intenses

à travers les époques (cf. supra 3.1.). Ces transformations sont en dernière instance plus importantes pour une langue que le simple inventaire de ses formes lexicales originelles, catégorisé d'après les langues d'origine. Mais il est infiniment plus difficile de structurer et de synthétiser le changement sémantique et dérivationnel que d'établir les couches étymologiques, même dans une langue aussi bien documentée et analysée que le français; à l'heure actuelle, des quantifications précises sont impossibles dans ce domaine, même si parfois des orientations générales se dessinent. Il est clair, notamment, que les mots latins héréditaires ont connu une plus grande activité non seulement dérivationnelle mais aussi sémantique (comme centre d'expansion et d'attraction) que les mots savants ou les autres emprunts, même au francique.

La meilleure manière de structurer les phénomènes de changement sémantique est de partir non pas des mots mais des concepts sous-jacents; ces derniers ne sont accessibles qu'à partir des sens lexicaux des mots mais ils permettent ensuite d'établir plus clairement des filiations et des développements dans le lexique (cf. infra 5., *L'interprétation onomasiologique*). La distinction entre sens lexical et concept, fréquente dans les études d'histoire littéraire ou culturelle, a permis de formuler des questions comme:

- Quels mots peut-on associer au concept de "LANGUE" (réponse: *langue*, *idiome*, *langage*, *variété* etc.)?
- Quels termes apparaissent pour évoquer des "SENTIMENTS" dans les chansons françaises du XIII^e s.?

Le modèle de référence pour cette approche cognitive et onomasiologique est le pentagone sémiotique, développé par Raible (1983) sur la base du trapèze sémiotique de Klaus Heger, élargi par Blank sous forme d'un carré sémiotique (1997) et nouvellement précisé par nous-même (2007; cf. aussi Lebsanft / Gleßgen 2004, 14–17); cf. notre représentation du carré sémiotique (→ fig. 237.1.).

Sur la base du 'modèle sémantique des trois niveaux' (Blank 1997), ce diagramme intègre dans le signe linguistique non seulement le sens lexical dénotatif («contenu sémémique»), mais aussi son «contenu syntagmatique» (valence verbale, collocations, phraséologismes), «relationnel» (polysémie, champs sémantiques, *frames*) et «diasystématique» (marques variationnelles). Dans

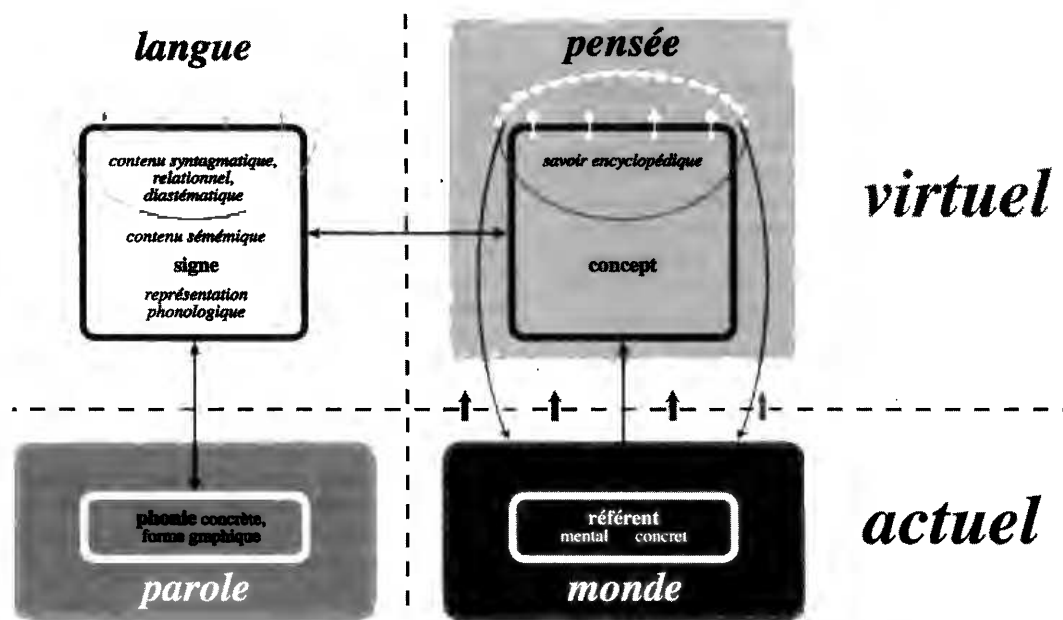


Fig. 237.1. Le carré sémiotique (d'après Gleßgen 2007, 239).

ces domaines, le signe linguistique s'ouvre vers les autres éléments constitutifs de la «langue», de la même manière que le concept s'ouvre vers la «pensée» à travers le «savoir encyclopédique»: une partie de ce savoir est immédiatement lié au concept (le 'savoir partagé' dans une communauté de locuteurs), une autre partie est plus individuelle et se trouve en interaction avec d'autres données dans la pensée personnelle.

La genèse d'un mot suppose dans cette logique différentes étapes:

(1) Les deux types de référents (mental comme la "BEAUTÉ", concret comme le "VOLCAN"), actualisés dans un énoncé, se construisent en interaction avec le «savoir encyclopédique»: un "VOLCAN" n'est distingué d'autres formations montagneuses qu'à partir du moment où nous définissons ses caractéristiques. Un référent potentiel devient donc référent en œuvre et «actuel», s'il fait objet d'une conceptualisation.

(2) La genèse d'un concept suppose un intérêt particulier pour un référent autour duquel se développe alors un savoir encyclopédique; le concept naît comme une agglomération de ce savoir encyclopédique, condensé dans des traits saillants, caractéristiques du référent et du concept en question (= catégorisation).

(3) Tout concept peut être exprimé par une paraphrase ou bien – dans le cas où il existe un intérêt très marqué pour un concept – par un terme

défini, créé pour cette fin (= genèse d'un mot); une telle 'verbalisation' utilise l'inventaire lexical existant, en attribuant un nouveau sens à une forme donnée ou modifiée par dérivation, en créant un phraséologisme ou en opérant un emprunt; le changement sémantique d'une forme provoque alors un effet de polysémie qui facilite en même temps la gestion du vocabulaire (cela demande un moindre effort cognitif de gérer 10.000 formes avec 30.000 sens que de gérer 30.000 formes). Un mot reflète donc le savoir partagé et considéré comme important par une communauté de locuteurs; il est lié à une langue particulière, contrairement au concept, non-linguistique, dont il prouve par ailleurs l'existence.

Lors d'une énonciation, une forme phonétique (ou graphique) évoque un sens lexical qui évoque à son tour un concept et ce dernier, à travers le référent, renvoie au savoir encyclopédique qui lui est associé.

La description des concepts verbalisés dans une langue met en relief l'intérêt particulier des hommes d'une époque et d'une culture donnée pour certains concepts et référents. L'histoire des évolutions sémantiques permet par conséquent des interprétations cognitives et historiques uniques. Par ailleurs, le changement sémantique permet de mieux comprendre la construction d'un sens linguistique et par là le fonctionnement de la langue, même en synchronie.

4.1.2. Les mécanismes du changement sémantique

La classification des mécanismes du changement sémantique connaît une longue tradition d'études, dans laquelle la romanistique a joué un rôle réel; mais il s'agit naturellement d'interrogations qui appartiennent dans leur visée à la linguistique générale. Les interprétations rhétoriques et psychologiques (cf. notamment Bréal, Michel, *Essai de sémantique*, Paris, 1897; Roudet, Léonce, *Sur la classification psychologique des changements sémantiques*, JPNp 18 (1921), 676–692; Ullmann, Stephen, *Précis de sémantique française*, Paris, 1952) ont été enrichies par la tradition structuraliste et nouvellement par la linguistique cognitive (cf. Wunderli 1990 et, surtout, les synthèses de Blank 1997; 2001; → art. 28; aussi → art. 259, 2.2.2.; Lebsanft / Gleßgen 2004). D'après la réflexion de Léonce Roudet (ib.), tout changement sémantique fait appel à la proximité («contiguïté») ou à la similitude entre le mot de départ et le nouveau mot, sachant que ces deux types peuvent s'appliquer aussi bien au sens des mots qu'à leur forme; cf. le schéma (→ fig. 237.2.).

	contiguïté	similarité
contenu / concept	métonymie	métaphore
forme / expression	ellipse	étymologie populaire

Fig. 237.2. Les mécanismes du changement sémantique d'après Roudet

Une contiguïté des deux concepts en question – sous-jacents aux deux sens – provoque une métonymie (par ex. *boire un verre* pour “boire le contenu d'un verre”), une similarité des deux concepts une métaphore (par ex. *trompe* pour “nez”). Une contiguïté des deux formes est à l'origine d'une ellipse (par ex. *diligence* dans le sens de “carrosse de diligence”, *cabinet* dans les sens de “cabinet d'avocat; cabinet de toilette”); une similarité des formes peut produire une étymologie populaire (par ex. *miniature* a le sens de “genre de peinture de petite dimension” par rapprochement avec la famille de lat. MINUS, même si étymologiquement il s'agit d'un dérivé de l'italianisme *minier* “peindre au minium” [1574], à son tour de MINIUM “oxyde de plomb, poudre de couleur rouge, minium”, couleur utilisée dans l'enluminure).

Tous les changements sémantiques peuvent être regroupés sous ces quatre catégories, avec un certain nombre de sous-catégories pour la métaphore et, notamment, la métonymie.

Définitions. – La métonymie exploite des contiguïtés ou des dépendances logiques entre deux concepts dans le monde de référence. Les contiguïtés les plus fréquemment actualisées dans le changement sémantique se placent à l'intérieur d'un champ notionnel (par ex. les dénominations des différents “ANIMAUX DE LA CAMPAGNE” qui donnent lieu à des substitutions sémantiques: lat. SÖREX “musaraigne” → “souris”, cf. pour tous les exemples de ce chapitre Blank 1997). D'autres relations métonymiques concernent le lien entre le tout et une de ses parties (par ex. *l'électeur* pour “l'ensemble des électeurs”) ou encore celui entre une cause et son effet (par ex. fr. *curieux* “qui éprouve de l'intérêt pour qqch” XII^e s. → “qui suscite un intérêt particulier” 1559 [= auto-converse, cf. infra]). Parallèlement, des contiguïtés peuvent s'instaurer entre des mots utilisés dans des contextes semblables (= frames; par ex. dans le *frame* “restaurant”: *nous avons quatre couverts de plus à la table 15* pour “quatre convives”).

La métaphore utilise au contraire le principe psychologique de la similarité des concepts qui appartiennent à des domaines notionnels et, pour la plupart, à des contextes d'usage différents; cf. l'exemple classique des dénominations métaphoriques de la “tête” dans les langues romanes qui partent de concepts initiaux différents évoquant la forme ronde de la tête (cf. Blank 1998; Pfister 2004):

- Le type le plus répandu TESTA veut dire en latin “pot; tesson” (> fr. *tête*, cf. aussi engad. / it. / sard. *testa*, aesp. *tiesta*, occ. *testa*; le changement sémantique mène d'abord par métaphore au sens “crâne” et ensuite par métonymie à celui de “tête”); cf. aussi les formes plus récentes fr.pop. *terriner*, *calebasse*, *carafe*, *cafetière*;
- des dénominations plus récentes reposent sur le sémantisme de “fruit rond ou gros” (fr. *poire* [se fendre la poire]; cf. all. *Kürbis*, *Birne*); d'autres langues (grec, persan, sanskrit) invoquent pour la tête le “sommet”, la “partie supérieure”.

Si la distinction théorique entre la métonymie et la métaphore est claire, ce n'est pas toujours facile de savoir dans un cas concret s'il s'agit d'une proximité ou d'une similarité. Cela suppose une connaissance précise des différentes étapes et contextes histori-

ques de l'évolution des sens d'un mot. Là encore, le français comme l'italien ou, dans une moindre mesure, l'espagnol permettent des évaluations plus sûres que presque toutes les autres langues du monde.

La métonymie est largement plus fréquente que la métaphore, même si cette dernière est plus spectaculaire et nettement mieux étudiée.

Principes sous-jacents dans les changements. – De nombreuses métaphores, notamment, ont un caractère euphémique ou, plus fréquemment, dysphémique qui augmente la charge expressive du mot en question; cf. dans la Romania l'exemple de la "tête" ou encore les métaphores de UNGULA "griffe" → "ongle" (cf. it. *unglia*, esp. *uña*, pg. *unha*, roum. *unghie*) ou de GAMBA → "jambe" (cf. it. *gamba*). L'augmentation de la charge sémantique qu'impliquent ces changements est toujours suivie d'un affaiblissement du sens, dès le moment où la motivation initiale n'est plus consciente: le mot *ongle* n'évoque plus en français le sens de "griffe" et perd alors son caractère d'expression imagée.

Le cheminement est analogue pour les métonymies mélioratives (euphémiques) de *mon-sieur*, *ma-demoiselle*, *ma-dame*: en français ancien, ces termes étaient réservés à la noblesse; leur utilisation pour des roturiers impliquait donc au début un élément mélioratif qui – par affaiblissement de sens (usure) – s'est estompé aujourd'hui.

Les métonymies et – encore plus – les métaphores font par ailleurs appel aux traits distinctifs qui constituent la hiérarchie des traits animés, profondément ancrée dans la structure mentale:

être humain – animal – plante – objet comptable (naturel / artisanal) – objet non comptable – abstrait.

Un trait saillant lié à une de ces catégories peut être métaphoriquement appliqué à une autre (par ex. *ruminer* appliqué à l'homme: le trait distinctif 'non humain' est remplacé par le trait 'humain', tout en gardant l'essentiel des autres éléments sémantiques du mot); de telles transpositions donnent lieu à de nombreuses métaphores occasionnelles; cf. quelques exemples lexicalisés:

- [+ humain] → [+ objet]: lat. *SATELLĒS* "garde du corps" → fr. etc. *satellite* "satellite"
- objet → humain: fr. *cafetière* "cafetière" → fr. pop. ~ "tête"
- animal → humain: fr. *renard* "renard" → "homme rusé" (les deux sens sont attestés au XIII^e s.)

- animal → objet: fr. *aile*, it. / esp. / pg. *ala* "aile" → "aile d'un bâtiment" (XII^e s. pour les deux sens); fr. *souris* etc. "souris" → "souris de l'ordinateur" (1983, calque de l'anglais)
- concret → abstrait: lat. *SAPERE* "avoir le goût de qqch" → rom. "comprendre, savoir" (on passe de "pouvoir distinguer les goûts" à "pouvoir distinguer les concepts")
- abstrait → concret: lat. *TROPARE* (IX^e s.) "parler en tropes, inventer des mélodies" → "faire des compositions" (= métonymie) → "trouver" (= métaphore); (cf. Pfister / Lupis 2001, 119–123; → art. 27, 2.3.).

La substitution de traits distinctifs est moins typique pour la métonymie, mais elle existe; cf.:

- humain → objet: fr. *courrier* "messenger; facteur" (XIV^e s.) → "courrier, lettre" (1770)
- objet → humain: fr. *cabinet* "lieu où les ministres tiennent conseil" (1631) → "ensemble des ministres" (1708; datations du FEW et du TLFi à revoir; la genèse des deux sens à partir de "pièce retirée" est chronologiquement proche).

Classement des métonymies d'après les catégories du champ notionnel. – Dans l'extrême majorité des cas, la contiguïté, caractéristique pour une métonymie, repose sur l'appartenance des deux notions au même champ notionnel. Dans une optique structuraliste, cela permet la distinction de (co-)hyponymes (*mouche*, *abeille*) et d'hyperonymes (*insecte*), souvent parallèle à l'optique cognitive d'une organisation des champs notionnels autour de prototypes; les deux vues introduisent les catégories logiques de l'élargissement ou de la restriction de sens: si *blé* (< *BLATUM*) passe du sens latin "céréales (en général)" au sens français de "froment" cette restriction de sens de l'hyperonyme à l'hyponyme (passage du genre / *GENUS* à l'espèce / *SPECIES*) utilise le principe du 'meilleur exemple'; de même pour l'élargissement de sens: si le mot lat. *PANARIUM* – "panier à pain" finit par désigner dans les langues romanes un "panier" (en général), c'est que le panier à pain devait être l'exemple le plus prototypique (= le plus fréquent?) pour un panier; cf. d'autres exemples pour une restriction de sens:

- lat. *VIVENDA* – "ce qui est nécessaire pour vivre" → esp. *vivanda* "qqch à manger", afr. *viande* "id." → fr. mod. *viande* "viande" (la viande n'était pas la nourriture la plus habituelle, mais la plus appréciée)
- lat. *HOMO* "homme (être humain)" → fr. *homme*, it. *uomo*, esp. *hombre* "id.; homme (être masculin)" (le changement suppose une société patriarcale; ce constat est confirmé par la

grammaticalisation ultérieure d'afr. *om* en fr. *on* "on", qui trouve un équivalent dans le synonyme all. *man* < *mann* "homme").

Dans le cas relativement fréquent d'une relation cohyponymique, deux concepts placés sous un même hyperonyme sont rapprochés ('glissement de sens'), cf. par ex. fr. *chapelle* → créole Réunion *sapèl* "temple indien".

Auto-converse. – L'auto-converse (ou 'métonymie interne') repose sur l'inversion des deux rôles sémantiques d'actant, de l'acteur et du patient. Le phénomène se produit surtout dans des verbes: fr. *louer* désigne par ex. autant l'action de "prendre en location" que celle de "donner en location" (de même it. *affittare*, esp. *alquilar*, cat. *arrendar*); cf. aussi la transformation déjà latine du lat. tard. *INODIARE* "haïr (qqn)" vers "être haïssable (pour qqn)" (→ fr. *ennuyer* "remplir d'ennui etc.", avec affaiblissement de sens). Il s'agit là de la lexicalisation d'une relation sémantique grammaticale. Les auto-converses sont moins caractéristiques pour des noms ou adjectifs (mais cf. par ex. lat. *HOSPES* "personne qui donne l'hospitalité" → "invité" et l'exemple donné supra fr. *intéressant*).

Classement des métonymies d'après le contexte d'usage. – En principe, toutes les métonymies peuvent être classées en relation avec un contexte d'usage (ou *frame*) défini. Blank (1997, 251s.) distinguait des changements où les deux concepts concernés sont présents dans un même cadre (= copésence) et d'autres, où ils entretiennent une relation de succession; ces deux types chevauchent les trois autres catégories des contiguïtés temporelles, spatiales et causales; cf. les exemples:

- contiguïté temporelle + succession (de deux laps de temps): lat. *VIGILIA* "veillée (avant une fête religieuse)" → fr. *veille* "jour qui en précède un autre", it. *vigilia* (= en même temps un élargissement de sens)
- contiguïté spatiale + coprésence; pour deux parties d'un objet (corps humain): lat. tard. *COXA* "hanche" → "cuisse", lat. *BUCCA* "joue" → fr. *bouche* "bouche", it. *bocca* (= cohyponymie); pour la partie d'un ensemble: fr. *âme* "âme" → "homme" (= type partie-tout, élargissement de sens)
- contiguïté causale + coprésence: afr. *message* "message" → "messenger" (= auto-converse: objet de l'action → acteur); lat. *TESTIMONIUM* "témoignage" → afr. *témoin* "témoin" (= auto-converse: action → acteur)
- contiguïté causale + succession (= cause-effet): fr. *addition* "opération consistant à réunir en un seul nombre toutes les unités dans plu-

sieurs autres" → "note présentant le total des dépenses effectuées au restaurant" (= élargissement de sens: action → effet de l'action); lat. *SUPERFLUUS* "surabondant, excessif" → fr. *superflu* "inutile" (= cohyponymie).

Sous-catégories de la métaphore. – La métaphore ne permet pas des classements aussi fins que la métonymie; elle est plus rare et – par l'appartenance des deux notions concernées à deux *frames* différents – plus imprévisible. Un type néanmoins bien identifiable est l'antiphrase (cf. ib., 220ss.); celle-ci met en œuvre une similarité qui repose sur un contraste très fort; c'est le cas, par ex., pour l'afr. *oste* "hôte, invite" qui désigne par la suite un "otage"; cf. aussi les dénominations antiphrastiques de la prostituée: fr. *respectueuse*, *religieuse* (avec des élargissements à l'intérieur du même *frame*, cf. *couvent* "maison close, maison de prostitution", (*mère*) *abbesse* "dirigeante d'une maison close"; cf. aussi it. *benefattrice* normalement "bienfaitrice", *serva di Dio* "serviteuse de Dieu"; pg. *filha-de-santo* "fille d'un saint").

L'auto-antonymie correspond à une opposition encore plus extrême entre deux pôles sémantiques, comme par ex. lat. chrét. *DEVOTARE* "invoquer en dévotion; dédier" → "maudire" (le changement sémantique s'est produit d'abord dans le modèle gr. *αναθέμα* "offrande votive" → gr. biblique "malédiction"; cf. FEW s.v. *ANATHEMA*); cf. aussi:

- fr. *rien* "quelque chose" – "rien", *aucun* "quelqu'un" – "personne", *personne* "id."
- lat. *SACER* "sacré" → fr. *sacré* "id." – "maudit" (*sacré menteur!*).

Ces deux types de changement sont très remarquables d'un point de vue sémantique – et s'expliquent soit par des faits de syntaxe soit par des évolutions socio-historiques – mais elles sont aussi peu fréquentes dans l'usage de la langue.

Les changements par motivation formelle. – Une motivation formelle pour un changement sémantique peut exploiter, elle-aussi, la similarité et ressemblance de deux formes (= étymologie populaire) ou leur contiguïté (= ellipse). Dans l'étymologie populaire, deux mots dont la forme se ressemblent sont interprétés comme appartenant à la même famille d'origine, ce qui provoque des interférences sémantiques, comme dans l'exemple donné de *miniature*; cf. aussi: afr. *faissier* "battre avec des joncs" (< *faisse* "faisceau de joncs" < *FASCIA*) est rapproché

de *fesse(s)* “derrière” (< FISSA “id.” < FIS-SUM “fendu, fissuré”) et voit par la suite sa forme et son sens altérés: *fesser* veut dire aujourd’hui “taper sur le derrière” (→ art. 28, 4.2.; 32, 4.2.).

L’étymologie populaire décrit donc le rattachement d’un mot à une autre famille génétique par fausse attribution étymologique; le nouveau rattachement provoque une réorganisation phono-sémantique des mots. L’étymologie populaire montre que l’homme raisonne continuellement sur la cohésion du vocabulaire même si son raisonnement reste inconscient; ce phénomène important pour la réflexion linguistique reste pourtant relativement rare dans l’évolution de la langue.

L’ellipse, bien plus fréquente, agit sur une forme en l’abrégeant; elle exploite la contiguïté entre une forme longue et une forme brève (cf. supra les ex. *diligence* et *cabinet*). Plus précisément, le changement sémantique ne concerne que la forme courte qui intègre, comme nouveau sens, celui du phraséologisme ou composé correspondant; cf. les exemples:

- fr. *bas-de-chausses* “chaussette” (1538) → *bas* (1552)
- lat. TEMPUS HIBERNUM “hiver” → fr. *hiver*, esp. *invierno*, it. / pg. *inverno*, roum. *iarnă*.

Les différents mécanismes dans l’histoire d’un mot. – Dans l’histoire concrète d’un mot, les différents mécanismes du changement sémantique interviennent souvent en série, tout en intégrant en même temps des transformations formelles; cf. l’exemple de l’histoire du mot fr. *grève*:

Le mot lat. parlé *GRAVA “gravier” survit en afr. *greve* avec le sens de “berge” (ca. 1140, par métonymie); à partir de ce terme se forme en divers lieux le toponyme *place de Grève* (de nouveau par métonymie: à Paris, c’est une place adjacente à la Seine, l’actuelle Place de l’Hôtel-de-Ville, anciennement le port de la ville); le toponyme parisien connaît une déonymisation sous la forme de *grève* “grève” (1805, ellipse et conversion contemporaines, mises en œuvre par une métonymie: sur cette place se réunissaient les chômeurs parisiens).

4.1.3. La portée interprétative des changements sémantiques

Les différents types du changement sémantique n’ont, a priori, rien de particulier en français; leur fréquence peut éventuellement varier selon le degré d’élaboration linguistique et les genres textuels: les journaux actuels produisent un nombre infini de métaphores

et de métonymies occasionnelles mais peu d’étymologies populaires. Mais la recherche n’est pas en mesure de préciser l’usage que fait une langue particulière de ces mécanismes dans une optique typologique.

La qualité du français réside toutefois dans l’avantage phénoménologique d’une bonne documentation historique, doublée de l’existence du FEW qui permettent des études systématiques et potentiellement quantificatrices des types du changement (cf. Lebsanft / Gleßgen 2004, 19–22, pour la quantification des changements latins-romans étudiés par Blank 1997).

Par ailleurs, l’étude des changements sémantiques fournit une base nécessaire pour les interprétations de type onomasiologique qui sont, à leur tour, plus spécifiques pour une langue à un moment historique donné (cf. infra 5.).

4.2. La formation des mots

4.2.1. Principes de formation

La transformation du vocabulaire par dérivation ou composition garantit en partie un renforcement du corps des mots abrégés par le changement phonétique (cf. l’exemple classique de l’afr. *af*, *e(f)* remplacé par *avette* ou l’occitanisme *abeille*). Mais elle ouvre surtout de nouvelles voies au changement sémantique par le changement de classes de mots (par ex. *réalis-er* + *-ation* → *réalisation*) et par la production de familles de mots (*jou-er*, *jou-eur*, *jou-euse*, *jou-et*, *jou-able*, *jou-jou*), sémantiquement diversifiées mais facilement gérables par le cerveau. Le sens d’un mot construit se nourrit dans une mesure très variable du sens de ses différents éléments constitutifs. Dans le cas le plus simple, le sens d’un suffixe est bien identifiable et s’intègre dans un amalgame sémantique nouveau et transparent avec le sens du mot de base; par ex.:

le $X_{\text{base verbale}} + \text{-eur}$ “celui qui X (de façon habituelle)”: *le vend-eur* “celui qui vend, celui dont la profession est de vendre”

la $X_{\text{base verbale}} + \text{-(t)-ion}$ “l’action de X et le résultat de cette action”, *la livr-aison* “l’action de livrer; ce qui est livré”.

Dans d’autres cas, tout aussi fréquents, le sens inféré par un affixe polysème ou peu spécifique produit un sens construit qui ne peut pas être dérivé immédiatement du sens de la base et du ‘programme’ de formation des mots; par ex.:

Le $X_{\text{base nominale}} + \text{-iste}$ "celui qui s'occupe professionnellement de X" a donné lieu à des sémantismes bien spécifiques: un *vendeur de journaux* s'occupe professionnellement avec des journaux, mais il n'est pas pour autant un *journal-iste*; un *paysag-iste* est un "peintre qui dessine des paysages" (ou "une personne qui élabore des plans de jardins"), alors qu'un *visag-iste* n'est pas un **"peintre de portraits"* (mais un "estéticien qui cherche à mettre en valeur le caractère du visage par la coiffure, le maquillage").

Le changement phonétique et lexical peut augmenter le décalage entre les anciens segments d'une forme construite et le dérivé; il accentue donc la coprésence, dans la langue, de mots construits transparents et opaques: certains mots construits anciens ne sont pas reconnaissables en synchronie actuelle comme tels (par ex. *soleil* < SOL + ICULU- ou *liesse* "(litt.) joie débordante et collective", isolé après la disparition d'afr. *lié* "gai" < LAETUS); d'autres ne peuvent plus être rattachés au mot de base d'origine (comme *vis-age* ou *pan-ier*, où seul le suffixe est identifiable en synchronie). Ici, comme dans d'autres cas, l'identification du morphème dérivationnel est plus facile que celle de la base, qui ne peut souvent être identifiée que grâce à la présence de ce premier. Dans un troisième groupe, le sens d'origine s'est perdu (un *luthier* construit des violons et non plus des luths; *en-gueuler* et *dé-gueuler* ne forment plus un couple d'antonymes). Mais il y a aussi des mots construits anciens qui restent transparents dans la formation et dans le sens malgré leur âge (par ex. *mari-age* [ca. 1135], *lang-age* [ca. 1160] et *pass-age* [ca. 1165] sont bien plus anciens que *régl-age* [ca. 1508] ou encore *bizut-age* [1949]). Ces différentes catégories se rangent dans un continuum, entre un pôle plutôt lexical (opaque) et un pôle plutôt grammatical (transparent sur la base d'un programme de formation de mots en vigueur, cf. Laca 1986: *Die Wortbildung als Grammatik des Wortschatzes*).

La situation est encore complexifiée, dans le cas du français et des autres langues romanes, par la présence de nombreux emprunts latins datant de toutes les époques, autant au niveau des bases qu'au niveau des affixes, et par la présence d'affixes latins qui restent reconnaissables et productifs même à l'époque romane. Décrire, de façon synthétique, les évolutions dérivationnelles à travers l'histoire du français n'est donc pas beaucoup plus simple que décrire les évolutions sé-

mantiques non-dérivationnelles (→ art. 259, 2.2.3.). Il est certain qu'il y a des époques de productivité forte pour certains affixes et certains types de formations; il existe aussi des accélérations dans certains domaines et époques: actuellement, par ex., le français 'périphérique' des jeunes et des banlieues en France ou celui de l'Afrique subsaharienne connaissent une dérivation plus productive que le français standard qui toutefois se livre à de nombreuses formations occasionnelles, notamment dans l'usage journalistique ou de la publicité.

4.2.2. Les mécanismes de dérivation en français

La dérivation par préfixe. – La préfixation simple modifie ou intensifie le sens lexical mais ne produit pas de changement de catégorie grammaticale. Toutes les langues romanes connaissent une présence forte des préfixes AD-, IN-, DE- / DIS-, EX-, RE-, CUM-, SUB-, SUPER-, TRANS-, EXTRA-, très fréquents en latin parlé tardif en raison de leur caractère d'intensification; par ex.:

AD-BATTUERE > ABBATT(U)ERE > fr. *abattre*, cf. it. *abbattere*, occ. cat. *abatre*, roum. *a abate* "détourner, distraire"

*EX-CAMBIARE > fr. *échanger*, cf. it. *scambiare*, occ. *escambiar*, roum. *a schimba*

Dans les deux cas, il s'agit d'une formation du latin (tardif) qui survit dans la plupart des langues romanes; les mots en question ne sont donc pas des dérivés romans.

La préfixation simple concerne surtout des verbes, bien qu'il existe aussi des dérivés nominaux (par ex. *nom* → *prénom*). Les bases nominales ou adjectivales produisent plutôt des préfixations qui impliquent un changement de catégorie grammaticale, ce qui suppose l'ajout d'un morphème flexionnel; cf.:

BRAC(C)HIU- → fr. *embrasser* (ca. 1100), cf. it. *abbracciare*

LONGU- → fr. *allonger* (ca. 1160), cf. roum. *a alun-ga* "chasser; pourchasser"

Il s'agit de formations parallèles mais indépendantes dans les différentes langues romanes, parfois à l'aide de préfixes différents (*em-brasser* vs. *ab-bracciare*); ceux-ci représentent toutefois des continuateurs des préfixes latins AD-, IN-, DE- / DIS- et non pas des préfixes propres à une langue romane.

Ces formations ont été considérées par la tradition romaniste pendant longtemps comme des formations parasynthétiques qui supposent l'addition combinée d'un préfixe et d'un suffixe; mais il s'agit plutôt de pré-

fixations entraînant un changement de flexion (conversion; de la même manière, la base lexicale *dépens-* fonctionne dans les deux classes grammaticales du nom [*dépense(s)*] et du verbe [*dépens-er, -era*]).

Par ailleurs, il existe aussi de véritables formations parasyntétiques comme:

in- + *coll-* + *-able* → *incollable* (1968, formation récente sur base verbale, donc 'véritablement' parasyntétique); un contre-exemple est fr. *désagréable* (1275/80) qui ne provient pas directement de *DIS-GRATUM-ABILIS mais d'une base déjà dérivée *agréable* (cf. it. *sgradevole*, esp. *desagradable*, pg. *desagradável*, sans doute des formations parallèles sur *gradevole* etc.).

La dérivation par suffixe. – La dérivation par suffixe est plus diversifiée dans l'histoire du français et généralement de la Romania; elle soulève par conséquent plus de problèmes formels et sémantiques que la dérivation préfixale. La suffixation entraîne souvent un changement de catégorie grammaticale (noms, noms propres, verbes, adjectifs → noms, (noms propres), verbes, adjectifs, adverbes) ou de genre (pour les noms). La complexité de ces formations ressort des exemples suivant pour des noms déverbaux; ces cas montrent aussi qu'un traitement historique satisfaisant des dérivés suppose la prise en considération non seulement du latin mais aussi des autres langues romanes pour établir l'époque de genèse et l'impact précis de la dérivation (cf. la liste de suffixes dans Brunet 1981, 503–598, et le traitement diachronique de chacun d'entre eux dans le TLF[j]):

-(T)URA

Formation latine: fr. *couverture* "pièce de tissu" (ca. 1180) < COOPERTURA ← COOPERIRE; cf. log. *kobertura* "toit"

Formation française / romane: fr. *dorure* (< id., *doreure*, ca. 1167) ← D(E)AURARE; cf. occ. / cat. *dauradura* "dorure" (XIV^e s.)

Commentaire: le suffixe -(T)URA, productif à l'époque latine, a connu des successeurs dans les différentes langues romanes qui sont restés productifs à leur tour (quant aux dates, il faut considérer que les mots attestés au XII^e s. se sont formés pour l'immense majorité d'entre eux déjà à une époque antérieure, avant l'apparition de l'écrit en roman); par ailleurs, la forme *couverture* est un emprunt au latin, contrairement à la forme héréditaire sarde.

D'un point de vue sémantique, les formations du type $X_{\text{base verbale}} + \text{-(T)URA}$ restent peu spécifiques "ce qui a trait d'une manière variable avec X (en fonction des référents ou du contexte d'utilisation)", ou encore "l'action ou le résultat d'une ac-

tion, désignée par X"; étant donnée la nature catégorielle du nom, il s'agit souvent des résultats d'une action (*dorure*) mais l'action peut rester présente; les différences sémantiques concrètes sont illustrées par l'opposition entre "toit" et "pièce de tissu".

-MENTU

Formation non latine mais néanmoins panromane: fr. *mouvement* (ca. 1100) ← MOVERE; cf. it. *movimento* (XIII^e s.), occ. *mo(u)vemen(t)* (XIII^e s.), cat. *moviment*, esp. *movimiento* (1250), pg. *movimento* (s. a.)

Commentaire: la formation peut a priori appartenir à l'époque très tardive du latin; elle pourrait aussi provenir, par emprunt, du latin médiéval; ou elle peut être l'effet de formations parallèles mais indépendantes dans les différentes langues romanes. Les différentes dates de première attestation illustrent une différence non pas au niveau de l'époque de la genèse des mots construits mais au niveau de la mise à l'écrit des langues romanes et de l'état de recherche respectif.

Les inférences sémantiques de -MENTU ne sont pas plus spécifiques que celles de -(T)URA.

-ANTIA

Cas semblable à -MENTU: fr. *espérance* "espoir" (ca. 1100); cf. it. *speranza*, occ. / pg. *esperança*, esp. *esperanza*.

-TOR

Formations romanes parallèles de type héréditaire: *pêcheur* (ca. 1140 [*pescheür*]); cf. it. *pescatore*, esp. *pescador*, occ. *pescaire*

Commentaire: Le sémantisme de $X_{\text{base verbale}} + \text{-(T)OR}$ correspond au sens général "celui qui X"; les actions en question peuvent être liées à un événement unique (*vainqueur*) ou à des habitudes (*buveur*); de là se produit une extension sémantique à des "personnes qui s'occupent professionnellement de X" (*pêcheur, entraîneur*; cf. supra -iste).

-TION/-SION-

Formations parallèles et latinisantes, sans doute avec des interdépendances entre les différentes langues romanes: fr. *réalisation* (1509); cf. it. *realizzazione*, esp. *realización*, occ. / cat. *realització*.

Les cas de figure présentés se répètent dans les autres types de formations qui impliquent une transformation catégorielle (cf. pour certains exemples Lüdtke 1996):

noms déadjectivaux:

-IA: fr. *courtoisie* (1155 [*curteisie*]); cf. occ. / cat. / pg. *cortesia*, esp. *cortesía* (formes semi-savantes; sémantisme peu spécifique "qualité d'être X")

-TAS / -TATE-: fr. *bonté* (ca. 1100); it. *bontà*, roum. *bunătațe*, esp. *bondad*, pg. *bondade*; occ. / cat. *falsedat*

-TITIA: fr. *justice* (ca. 1050, forme savante), *tristesse* (ca. 1165 [*tristece*], ca. 1180 [*tristesse*], forme semi-savante), *franchise* (ca. 1130, forme héréditaire); cf. occ. / cat. *tristesa*, esp. / pg. *tristeza*

adjectifs déverbaux:

-ABILE: *faisable* (ca. 1350), *jouable* (1741); cf. TLFi -able / -ible / -uble

-ATORIU: *diffamatoire* (1380)

-(r)IVU: *abusif* (1371) [terme grammatical], *qualificatif* (1756); mots savants

adjectifs dénominaux:

-ALE: *musical* (XV^e s.); cf. TLFi -al / -el

-ANU: *australien* (1831, DI s.v. *Australiä*)

-IBILE: *paisible* (déb. XII^e s., ← *paix*)

-ICU: *volcanique* (1778, ← *volcan*)

-ISCU: *gargantuesque* (1707, ← *Gargantua*) [exemple d'une déonymisation]

verbes dénominaux:

-IDIARE: fr. *guerroyer* (ca. 1100 [guerreier])

-IZARE: fr. *tyranniser* (1365)

adverbes déadjectivaux:

Formations en -MENTE (adj. au fém.): *lentement*, *antérieurement*, *bonnement* (cf. TLFi -ment²)

changement de genre / 'motion' (masc. → fém.):

-U(S) / -A: par ex. -ARIU / -A: *laitier* / -ière; cf. it. *lat-taio* / -a, esp. *lechero* / -a, pg. *leitero* / -a

-TOR / -TRIX: *moteur* – *motrice*; cf. occ. *trobador* – *trobairitz*

-ISSA: *maîtresse*

Dans ce domaine morphologique s'inscrit l'actuelle discussion sur les dénominations féminines de professions ou de fonctions (par ex. *le / la ministre*, *avocat / avocate*, *docteur / doctoresse*, *professeur / professeure*).

Les transformations sémantiques sont moins développées dans les suffixations surtout nominales sans changement catégoriel; cf. les exemples:

diminutifs (avec différents sens secondaires):

-ELLA / -ELLU: *mademoiselle* (1471), *damoiseau* (ca. 1135 [dameise])

-ITTA / -ITTU: afr. *estoitete*, fr.mod. *jouet* (mil. XIII^e s. [jüez]), *coffret* (déb. XIV^e s.), *archet* (XI-V^e s. [ca. 1160 terme arch.]), où le suffixe est aujourd'hui pétrifié; mais: *maisonnette* (1176/81), avec transparence du suffixe en fr. contemporain; cf. TLFi -et / -ette.

noms relationnels (= noms qui ont la fonction "X en relation avec le mot de base"), faisant référence à des personnes:

-ARIU: *charbonnier* (fin XII^e s.); cf. it. *carbonaio*, *carbonaro*, roum. *cărbunar*, cat. *carboner*, esp. *carbonero*, pg. *carvoeiro*; savant: *fonctionnaire* (1770), *funzionario* etc.; [cf. -ND-ARIU: *lavandière* (1180), *filandière* (1292)].

noms relationnels, faisant référence à des objets:

-ARIU: *grenier*, *guépier*.

collectifs:

-ETU: afr. *sapoi* "forêt de sapins" (formations fréquentes dans des toponymes comme *Aulnois* < *aln-etum* ou *Rouvray* < **robur-etum*, → art. 239, 2.2.4.)

-ATICU > gallorom. -age: *visage*, *langage*, *feuillage*, afr. *barnage* "ensemble des hommes libres", cf. pg. -agem (*folhagem*), it. -aggio (*coraggio*)

Le suffixe -ATICU sert parallèlement non pas comme collectif mais comme élément de formation de déverbaux qui correspondent à des noms d'action (fr. *lavage*, *mariage*, *passage*; cf. les exemples datés supra 3.2. et TLFi -age: «Les créations nouv. en -age sont innombrables»).

-ALIA: *ferraille* (ca. 1349), cf. it. *ferraglia*.

Dans les formes adjectivales, les suffixations peuvent indiquer une approximation (par ex. *maladif*, *vieillot*, *ellipsoïde*).

Mécanismes particuliers de la dérivation. – A côté des deux grands mécanismes de préfixation et de suffixation s'instaurent d'autres types de formations qui peuvent être regroupés sous les principes de la dérivation. Selon les époques ou les variétés en question, ils sont plus ou moins productifs; mais là encore, le français partage l'essentiel de ces mécanismes particuliers avec d'autres langues romanes:

– Conversion: passage d'une classe de mots à une autre par simple changement du morphème flexionnel (par ex.: *sucré* → *sucrer* ou – à l'inverse – *donner* → (*une*) *donne*). Le cheminement d'une conversion ne se comprend souvent qu'en diachronie; les locuteurs en synchronie ne reconnaissent pas toujours quelle est la base et quel est le dérivé.

– Dérivation impropre ou dérivation zéro: une conversion sans modification formelle; c'est le cas, par ex., de la nominalisation d'un infinitif, fréquente en ancien français et en roumain mais largement attestée aussi dans les autres langues romanes.

Ce type de conversion intervient aussi dans la déonymisation (à côté de la suffixation et de la formation de phraséologismes, cf. Schweickard 1992); cf. les types "nom de l'inventeur ou du propagateur" → "nom de l'invention" (*zeppelin* 1907, *poubelle* 1890) ou encore "nom de marque" → "produit en général" (*kleenex* "mouchoir en papier" 1965).

Une transformation semblable, mais sémantiquement plus drastique est l'antonomase qui désigne autant les passages des noms propres vers les noms communs que l'inverse; la motivation repose alors sur des traits saillants: dans la formation d'un nom propre, un trait de caractère qui désigne ensuite tout le personnage (par ex. *le Seigneur* pour "Dieu"), dans la déonymisation, un individu qui représente de façon typique une catégorie (par ex. *un Tartufe* "hypocrite" 1665 [*Tartuphe*]); le passage du nom propre vers le nom commun n'est plus, ici, une métonymie, mais une métaphore (un *Tartufe* ressemble au personnage théâtral alors qu'un *kleenex* est vraiment un mouchoir en papier).

– **Rétroformation ou dérivation régressive:** mots simples formés sur des dérivés; par ex. fr. *diplomatique* adj. (1721) → *diplomatie* n.f. (1790), *diplomate* n.m. (1792).

– **Affixoïdes:** dérivation typique du XX^e s. où des morphèmes lexicaux pouvant fonctionner comme des lexèmes libres sont utilisés en série comme des affixes (par ex.: *auto-* (*suggestion, suggestionne, suggestion*), *super-*, *télé-*); l'utilisation de ces formes comme lexèmes libres (*super*) est normalement postérieure à leur emploi fréquent en tant qu'affixes.

– **Réduplication:** le type *fi-fille, mé-mère, jou-jou*; la reduplication met en œuvre une expressivité ou intensification linguistique; elle est souvent caractéristique du langage enfantin et des emprunts à celui-ci dans le langage familier.

– **Dérivation sans base:** une forme affixée ne suppose pas nécessairement, au moment de sa formation, une forme de base sans affixe; des termes comme les formes suffixées *mensuel* et *nuptial* sont entrés dans la langue en tant que latinismes, sans que les bases théoriques du type **mensu-* ou **nupt-* aient existé. De telles formations ne sont pas rares et ont souvent perturbé la reconstitution de la trajectoire historique de certaines familles de mots.

Par ailleurs, des bases existantes peuvent être altérées lors de la formation, comme *bijou|tlier*, *congol|lais* ou *voul|oyer* (où la base est élargie afin d'éviter le hiatus entre les deux morphèmes) ou comme *canadien* et *mot|lard* (où la base est abrégée). Ce phénomène est à distinguer du cas déjà mentionné, où une base donnée disparaît ou devient méconnaissable (cf. par ex. en fr. actuel *bouv|ier*, *pan|ier*, *pel|age* qui ne permettent plus de reconnaître les formes des bases latines BOV(IS), PAN(IS), PEL(L)(IS), même si les suffixes, eux, sont évidents; ex. de Grevisse ¹³1993, § 166 b 1^o). Dans de nombreux cas, des mots dérivés représentent des emprunts au latin, alors que la base correspondante a connu une évolution héréditaire ou semi-savante (par ex. *marin* vs. *mer*; *populaire* vs. *peuple* [semi-savant]); la relation entre le lexème libre et la base de la dérivation (identiques en latin) est alors rompue à l'époque des langues romanes; parfois, les emprunts viennent même suppléer la base (par ex. *oral* vs. *bouche*).

– **Dérivation sur un composé:** *conseil|ler municipal*. Ces formations possèdent un statut sémantique ambigu, puisque le premier terme est une forme apocopée d'un autre syntagme: *conseiller* n'est, en effet, pas "municipal", mais il s'agit d'un "membre du conseil municipal". Cette contradiction est fortement ressentie par les locuteurs, ce qui réduit l'action de cette formation.

Les divers exemples donnés soulignent la grande particularité de la dérivation: c'est sans doute le domaine lexical qui montre le plus grand nombre d'interférences entre les

différentes langues romanes (cf. Ernst 1995; → art. 258, 2.1.).

4.2.3. La composition en français

La composition est l'union de deux mots autonomes (= deux morphèmes lexicaux libres). Les liens sémantiques entre ces deux lexèmes sont très divers; il est difficile, dans la plupart des cas, de formuler des règles sémantiques qui dépassent la définition des deux éléments formateurs respectifs (comme "un X qui a trait d'une manière ou d'une autre à un Y"); le sens du composé dépend alors fortement des référents et du contexte d'usage des mots composés.

Il existe dans les langues romanes certains types de composés qui sont écrits aujourd'hui en un mot (*tournesol*) ou, plus fréquemment, avec tiret (*mot-clé*), d'autres qui contiennent plusieurs mots graphiques (*pommes de terre*). Dans ce dernier cas, il faut bien distinguer les syntagmes libres (par ex. *vêtement blanc*) des composés lexicalisés (*vin blanc, escalier roulant*): l'unité lexicale des composés peut être démontrée par l'impossibilité de décomposer l'ensemble sans transformer son sens (par ex. **Cette pomme est de terre, *Des pommes jaunes de terre, *Des pommes de terre sablonneuse*: l'attribution se réfère à l'ensemble du lexème composé).

Le type de composition le plus caractéristique est sans doute la composition nominale, bien moins fréquente dans les langues romanes que, par ex., dans les langues germaniques; ce mécanisme a néanmoins gagné de l'importance en français depuis le XX^e s.; cf. les exemples:

– **composition nominale déterminative:** un élément du composé détermine le sens de l'autre: *homme-grenouille* (1960), *mot-clé* (1971?), *assurance maladie* (sans tiret comme la plupart des formations récentes, très difficiles à dater par ailleurs, parce que la lexicographie ne les reconnaît pas comme composés), *poids mouche* (1936?). Dans les langues romanes, le premier élément est déterminé par le second (ordre déterminé-déterminant), contrairement aux langues germaniques (ordre déterminant-déterminé: *Frosch-mann, Schlüssel-wort, key-word*).

– **composition nominale cumulative:** le sens du composé est constitué par celui de ses deux éléments constitutifs de façon égale: *fille-mère* (1848), *moissonneuse-batteuse* (1931), *porte-fenêtre* (1932).

La formation de composés adjectivaux répond, elle aussi, à deux sémantismes différents, cumulatif ou intermédiaire:

- composition adj. + adj. de type cumulatif / additif: *sourd-muet* ([1564?] 1791)
- composition adj. + adj. avec le sens de “entre X et Y”: *franco-allemand* (1885).

Par ailleurs, la composition englobe la jonction de deux catégories grammaticales différentes (v. / prép. / adj. + n., v. + pron.); cf. les exemples souvent plus anciens:

- v. + n.: *tournesol* (1360), *couvre-chef* (XII^e s.), *tire-bouchon* (1718)
- v. + pron.: *fainéant* (1306 [v. 1200]), *vaurien* (1558)
- prép. + n.: fr. *survêtement* (1939 [1606]), *sous-pull* (1970)
- adj. + n.: *gentilhomme* (XIII^e s.), *haute tension* (1868).

Le type de composition le plus habituel en français comme dans la plupart des langues romanes est toutefois une formation de type syntagmatique. Celle-ci implique fréquemment des noms comme dans *pommes de terre* ou *machine à laver*; dans la structure de surface, ces formes sont identiques à des phraséologismes (cf. infra 4.3.) desquels elles se distinguent par la plus grande fixité (impossibilité de décomposer sans altération sémantique), par une plus grande fréquence dans l’usage et par la conscience des locuteurs qui les perçoivent comme des unités. La séparation reste pourtant difficile, d’autant plus si des formes sont parfaitement transparentes et productives à un moment historique donné (comme *rédacteur en chef* ou encore des composés n. + adj. comme *parking souterrain*). Des formations comme *tout à l’heure* ou *tout à fait* sont généralement considérées dans la lexicographie comme des phraséologismes adverbiaux mais ce sont des formes très stables, non décomposables, avec une transparence réduite et fonctionnant comme des unités dans la phrase. L’étude historique des composés doit donc prendre en considération la phraséologie et elle gagnerait sans doute à élargir le concept de composition.

4.2.4. Autres mécanismes de formation de mots

Croisements de mots. – Les croisements représentent un type de (trans-)formation ancien, bien que peu répandu. Une telle interférence se produit notamment lorsque la forme et le sens de deux mots sont proches; par ex.:

- le mot *trouble* semble provenir de *TURBULUS résultat d’un croisement entre TURBIDUS et TURBULENTUS

- dans le mot *craindre*, on suppose une interférence entre lat. *TREMULARE (< TREMERE) et un verbe celtique sémantiquement proche (cf. bret. *kriedien* “fait de trembler”).

De tels croisements se rapprochent de l’étymologie populaire (cf. encore le phraséologisme *il vaut mieux* qui suppose à l’origine un type **il faut mieux*). Les interférences peuvent aussi être conscientes ou volontaires comme cela est le cas dans les ‘mots-valises’, répandus depuis le XX^e s. (par ex. en fr. *auto|bus* 1906, *fran|glais* 1959 ou *res-to|route* 1954).

Délocutifs. – Les délocutifs transforment un élément du discours en un lexème. Ils interviennent fréquemment dans des formes avec une forte charge pragmatique, qui favorise l’évocation d’un contexte d’utilisation spécifique. Les bases verbales dominent dans ce mécanisme (par ex. *je-m’en-foutisme* 1891, (*un*) *m’as-tu-vu* v. 1800).

Mais il existe aussi de délocutifs basés sur des pronoms (cf. *vouvoyer* 1834) ou sur des actes d’énonciation comme les formules de salutation (*adieu* n.m. XII^e s.); cf. aussi les latinismes *factotum* (1552) ou *fac-similé* (1821 [1796]).

Abréviations et sigles. – Les abréviations et – encore plus – la formation de sigles ou acronymes sont des mécanismes de formation relativement récents et extrêmement productifs depuis la deuxième moitié du XX^e s. L’abréviation s’opère essentiellement par apocope (parmi les formes les plus anciennes: *métro|politain* 1891, *vélo|cipède* 1890, *prof|esseur* 1890; cf. aussi *impec|cable* 1950); son action par des formes occasionnelles s’est encore intensifiée au début du XXI^e s., notamment dans le langage des jeunes.

La formation de sigles ou acronymes caractérise l’essentiel des langues standard modernes, avec une productivité infinie depuis quelques décennies (par ex. UFR < *uni-té de formation et de recherche* pour l’ancienne “faculté universitaire”, D.E.U.G. /dœg ~ 'dœg/ < *Diplôme d’études universitaires générales*); les formes restent pourtant peu variantes et donnent rarement lieu à des dérivés comme C.A.P.E.S. /ka'pes/ (< *Certificat d’aptitude au professorat de l’enseignement supérieur*) → *capésien*.

Défigurations volontaires. – Les défigurations formelles volontaires représentent le type le plus excentrique dans la formation des mots. La transformation vise dans ces cas à camoufler la base de départ, ce qui est unique parmi les mécanismes de transforma-

tion lexicale. De telles défigurations sont plus ou moins conscientes pour les locuteurs et peuvent s'inscrire dans une volonté ludique.

Dans l'histoire ancienne du français, il s'agit essentiellement d'euphémismes, par ex. les *nomina sacra* du type *parbleu* (1540, ← *pardieu* < *par Dieu*!) pour éviter le nom de Dieu dans un juron. Plus récemment, une généralisation de ce mécanisme de défiguration a produit le *verlan* (1970 [1953] ← (*à l'envers* /lâ'vɛʁ/): un inventaire lexical qui repose sur l'inversion de certaines syllabes et qui a ses racines comme jargon argotique dans les années 1920 et 1930; le *verlan* s'est nouvellement développé depuis les années 1980, notamment dans le cadre du langage des banlieues et des jeunes (→ art. 206, 6.1.3.; cf. les ex. *laisse tomber* → *laisse béton* 1970, *Arabe(s)* → *Beur(s)* 1980 → *rebeu*, *femme* → *meuf* 1981, *pourri* → *ripou* 1985).

4.3. Contexte syntagmatique et phraséologie

La formation d'unités plurilexématiques constitue le troisième mécanisme interne pour transformer ou élargir le lexique. Le classement de telles unités est extrêmement épineux, d'autant plus que certaines d'entre elles peuvent être considérées comme des composés (cf. supra 4.2.3.). Les liens syntagmatiques entre les mots sont multiformes et s'inscrivent dans un continuum avec une intensité d'interaction croissante entre différents mots. Ces relations d'interaction privilégiées sont importantes pour structurer un discours; la plupart des lexèmes sont utilisés dans un nombre relativement limité de contextes habituels. Pour leur structuration, il est possible de distinguer trois paliers ni étanches ni uniformes sur un continuum:

(1) La valence verbale génère un premier niveau de liens syntagmatiques lexicalisés; elle concerne surtout les cadres valentiels des verbes (*jouer de* (**sur*) *la flûte*, *au foot*, *aider sa voisine à* (**pour*) *comprendre le cours*). Le nombre d'actants admis par les verbes et les prépositions qu'ils régissent représentent des restrictions d'ordre syntaxique, liées en même temps à un verbe défini. Ce premier palier se trouve par conséquent à cheval entre la grammaire et le lexique; d'un point de vue lexical, la forte individualité de certains verbes mériterait des études thématiques, d'autant plus dans l'optique de leur – notable – variation dans le temps et dans les genres textuels. Selon le sémantisme et l'ancrage contextuel d'un verbe, la 'portée' de celui-ci préfigure au delà des deux ou

trois actants la gestion d'autres compléments, notamment circonstantiels. Par ailleurs, les verbes opèrent souvent des restrictions solides sur le choix des termes concrets qui peuvent les accompagner. C'est un domaine d'étude en expansion et riche en perspectives (cf. pour le français contemporain les travaux de Gaston Gross, Peter Blumenthal, Achim Stein, Rolf Kailuweit; pour l'histoire ceux de Lene Schøsler).

(2) Les collocations correspondent à des liaisons plus spécifiques entre certains mots dans l'usage de la langue (par ex., l'on dira facilement que *l'eau est bonne*, non pas *agréable* ou *à température idéale*, ce qui, sémantiquement et syntaxiquement, serait tout aussi cohérent; cf. aussi *faire le naïf*, et non *jouer le naïf*, *prononcer un discours*, non pas *dire*, *énoncer* ou *présenter un discours*). L'utilisation de ces mots dans un même contexte syntagmatique est habituelle, mais les mots gardent néanmoins l'autonomie d'être construits avec d'autres termes. Les collocations se placent donc au niveau de la norme d'usage; elles se caractérisent surtout par une certaine fréquence et fixité; mais ces usages préférentiels ne représentent pas nécessairement une forte unité sémantique; par ailleurs, le sens de la collocation est directement compréhensible à partir des éléments qui la forment.

(3) Les phraséologismes ou idiomes indiquent des liaisons très fréquentes et lexicalisées entre différents mots et montrent une stabilité relative; de telles liaisons peuvent impliquer toutes les catégories grammaticales et entrent souvent dans des constructions syntaxiques inhabituelles en synchronie; leur sens d'ensemble ne peut pas être déduit facilement à partir des différents éléments de base (par ex. *le jeu n'en vaut pas la chandelle* "cela ne vaut pas la peine", *en mourir d'envie*). Ils ont les caractéristiques d'un sens relativement unitaire, d'une certaine fixité et – en même temps – variabilité. Leur non-transparence est en effet purement historique: au moment où un phraséologisme est créé, il est toujours transparent pour les utilisateurs. Mais une fois figé, les motivations de départ peuvent disparaître dans le monde des références ou dans la conscience des locuteurs (par ex. dans les théâtres modernes, l'éclairage est garanti par l'électricité et non pas par des bougies – *chandelles* –; par ailleurs, les pièces ne s'appellent plus *jeux*, ce qui rend incompréhensibles les origines de l'idiome cité).

Il est très difficile de distinguer un phraséologisme stable d'un composé, comme par ex. pour *maison close* "maison de prostitution", *bonne chance*, *à la bonne heure*, *à l'intérieur de*, *franc du collier*.

Dans l'évolution de la langue, les interactions habituelles entre les mots sont soumises à des changements tout comme les autres domaines de la langue, éventuellement avec des rythmes d'innovation supérieurs et avec une valeur symptomatique particulièrement forte pour des modes culturelles. C'est un chapitre jusqu'ici fort peu étudié en diachronie. C'est seulement pour les tournures idiomatiques qu'existent de nombreux répertoires de qualité inégale et quelques études comparées ou historiques ponctuelles (par ex. Schweickard 1990 ou différents articles de Gilles Roques; cf. aussi Lebsanft 1987 pour les formules de salutation en ancien français).

5. L'interprétation onomasiologique de l'histoire du vocabulaire français

Les différents mécanismes évolutifs (chap. 4) qui s'appliquent à un ensemble de lexèmes constitué (chap. 3) produisent une physionomie évolutive bien caractéristique en français comme pour toute autre langue. Puisque chaque mot connaît en dernière instance sa propre histoire, il ne sera sans doute jamais possible de rendre compte de façon synthétique d'une telle physionomie dans une optique typologique. C'est la particularité propre au lexique qui lui a valu à tort le jugement d'un domaine profondément hostile à la théorisation. Il n'en reste pas moins que si certaines questions interprétatives peuvent être traitées, elles resteront toujours partielles (combien de mots latins avec ou sans dérivation, avec ou sans changement sémantique, survivent à une époque donnée en français? combien de latinismes ou d'autres emprunts s'y sont ajoutés? etc.).

Les études d'onomasiologie historique permettent toutefois de considérer les mécanismes évolutifs et l'origine des mots dans une optique d'ensemble; elles se concentrent sur l'évolution de certains concepts, groupes de concepts ou domaines sémantiques plus larges, éventuellement ciblés dans certains genres textuels ou segments du diasystème, et prennent en considération le contexte socio-historique en question.

L'idée de départ de telles analyses est d'une grande simplicité; il s'agit d'identifier comment (1) pour un même concept (par ex.

la "tête") les mots changent dans le temps et sont remplacés par d'autres (par ex. *CAPUT* par *TESTA*); par la suite, elle peut utiliser les renseignements liés aux mots (sens, fréquence, contexte) pour établir (2) si le concept de départ s'est éventuellement transformé dans le temps (par ex. le remplacement de *ost* par *armée* au XV^e s. coïncide avec un moment où la nature des armées changeait). Dans le cas le plus simple, le concept reste inaltéré, par ex. l'évolution des mots désignant le concept "bouche":

Le mot lat. *os* "bouche" est remplacé par des mots signifiant "joue" (rapprochement métonymique, lié à l'idée que la "bouche" sert à *manger*) ou "gorge" (métonymie qui met en avant l'idée de *crier* / *produire des sons forts* / *parler*):

BUCCA "joue" → fr. *bouche*, cf. it. *bocca*; GALTA "joue" → béarn. *gaute*

GULA "gorge" → roum. *gură* (le fr.pop. *gueule* reflète une autre évolution, même s'il remonte aussi à l'étymon GULA: il naît par remplacement du trait "animal" par le trait "humain" et implique un élément dysphémique comme, à l'origine, *ongle* ou *griffe*).

Même si le concept n'a pas connu de transformations, les remplacements montrent que le concept "bouche" est associé avec les deux idées de *manger* et de *parler*, qui constitueraient alors les éléments saillants de ce concept.

La méthodologie onomasiologique a été développée à partir des travaux de Ernst Tappolet (sur les dénominations de la parenté dans la Romania: *Die alemannischen Lehnwörter in den Mundarten der französischen Schweiz*, Heidelberg, 1895) et de Adolf Zauner (sur les dénominations des parties du corps: *Die romanischen Namen der Körperteile. Eine onomasiologische Studie*, RF 14 [1903], 339–530) surtout dans la première moitié du XX^e s., en réunissant des mots qui désignent un concept donné ou plusieurs concepts voisins (cf. la synthèse bibliographique de Quadri 1952). Elle a mené à la proposition de plusieurs systèmes onomasiologiques globaux, notamment en romanistique, le *Begriffssystem* (R. Hallig / W. v. Wartburg, *Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas*, Berlin 1963 [1952]), dont l'utilité reste intacte encore aujourd'hui. Le courant des "mots et des choses" a été réinterprété par les théories sémantiques structurales d'abord (cf. Bernard Pottier, *Recherches sur l'analyse sémantique en linguistique et en traduction mécanique*, Nancy, 1963), par les

théories cognitives ensuite; ces dernières ont introduit le concept du prototype comme principe d'organisation au niveau cérébral (cf. Kleiber 1990) et mis en avant l'importance du niveau sémantique de base comme reflet premier du monde des références quotidiennes qui structurent notre univers mental (cf. Geeraerts 2004; cf. aussi l'aperçu de Koch 2001); elles ont aussi nouvellement fondé l'importance des traits distinctifs par la hiérarchie des traits animés (humain, animé, objet, abstrait, cf. supra 4.1.2.).

Les études de type onomasiologique – toutes théorisations confondues – permettent d'entrevoir des éléments de la structure mentale de la langue et de son interface avec la pensée et avec la culture: un changement sémantique reflète des associations sémantiques récurrentes ou stéréotypées dans la communauté de locuteurs en question. Les finalités de telles études sont de rendre visible des associations mentales de type universel – le réseau conceptuel est sans doute organisé de façon anthropocentrique, à partir du corps humain et des références quotidiennes qui nous entourent (la parenté, les animaux, la nourriture, les lieux d'habitation, la nature et les phénomènes climatiques, les professions etc.) – ou encore des concepts et évolutions historiquement contingentes, caractéristiques pour une culture donnée (cf. la mise en relief des transformations lexicales entre 1800 et 1950 de Brunet 1981, 688–708); l'onomasiologie fournit donc un point de départ pour des études en anthropologie historique, interprétable autant d'un point de vue culturel que d'un point de vue universel.

Les interrogations onomasiologiques concrètes sont toujours appliquées à des découpes bien définies parmi les langues ou les époques: elles peuvent englober toute une langue historique (par ex. le français du IX^e au XX^e s. comme Stefenelli 1981), ou se restreindre à une seule variété de langue (le dialecte lorrain, le français populaire) ou encore à un genre textuel bien défini (comme par ex. les textes littéraires); elles peuvent opérer diverses coupes chronologiques (le XVI^e s. comme siècle de transition entre l'ancienne langue et la langue moderne, l'époque des premiers textes, l'époque contemporaine) et englober plusieurs langues, romanes ou non romanes. Toutes les combinaisons sont possibles et a priori légitimes – et toutes ont recours aux mêmes techniques de base.

De nombreuses études spécifiques ont étudié l'évolution des mots désignant les

concepts les plus présents dans une langue – concernant les références quotidiennes –, ou encore des concepts clés pour la culture occidentale (cf. Schmitt 2001); le point de départ peut être alors onomasiologique, en partant d'un concept (comme par ex. Jean Renson, *Les dénominations du visage en français et dans les autres langues romanes. Etude sémantique et onomasiologique*, Paris, 2 vol., 1962), ou sémasiologique, en partant des mots donnés dont l'évolution est interprétée toutefois dans une optique onomasiologique (comme par ex. Hans Rheinfelder, *Das Wort 'Persona'*, Halle, 1928; cf. pour les divers travaux et leur méthodologie → art. 28, 2.).

Par ailleurs dominent des études sur des concepts littéraires, notamment au MA, ou sur le vocabulaire scientifique ou technique à différentes époques. Cf. deux exemples pour des études ciblées de concepts culturels de base:

– "argent": le vocabulaire médiéval fait ressortir le lien entre la possession d'animaux domestiques et l'argent dans une société pastorale; fr. *gagner*, it. *guadagnare* proviennent de a.b.frq. **waiðanjan* "paître, mener les bêtes aux champs", lat. *PECUNIA* provient de *PECUS* "animal domestique; mouton", l'angl. *fee* "paie" en dernière instance de l'a.b.frq. **fehu* "bétail" (cf. FEW s.v.)

– "travail": le lien entre le travail et la souffrance est établi par les termes romans fr. *travail*, esp. *trabajo* etc. < lat. *TRIPALIUM* "instrument de torture consistant en trois pieux auquel la victime était attachée"; les étymologies du mot lat. *LABOR* et du mot germ. *arbeit* reposent de la même manière sur des termes désignant la peine ou la souffrance (cf. Pfister / Lupis 2001, 115).

La méthodologie onomasiologique s'avère également très utile pour déterminer les raisons de la disparition d'un mot. Si l'introduction de nouveaux termes peut s'expliquer par des volontés de mise en scène de la part d'individus ou de groupes sociaux ou encore par des transformations extralinguistiques, la disparition régulière d'une partie du vocabulaire en est une contre-partie nécessaire; mais les règles qui décident des mots qui disparaîtront restent vagues et aléatoires; certains facteurs peuvent toutefois expliquer le remplacement d'un mot par un autre:

– une faiblesse phonétique marquée: *APEM* > afr. *af* ~ *ef* ~ *e* → *avette* ~ *abeille* (cf. supra; afr. *ost* → *armée* ([ɔst] serait réduit à [ɔ]); ici s'ajoutent comme motif extra-linguistique les transformations militaires déjà mentionnées)

- une relative faiblesse sémantique dans le cas de concepts qui appellent naturellement de l'expressivité: *IOCARE* "blaguer" (se maintient sémantiquement affaibli avec le sens de "jouer"), remplacé par fr. *blague*, *rigoler* etc.; it. *scherzo* (< germ.), esp. *burla* (origine inconnue)
- l'homophonie avec des mots, qui appartiennent au même contexte communicatif et peuvent donc être confondus (cf. Blank 1997, 354ss.):
- = afr. *moudre* "traire" (< *MULGERE*) ~ *moudre* "moudre" (< *MOLERE*)
- = afr. *noer* "nager" (< **NOTARE*) ~ *noyer*, *neïer* "noyer" (< *NECARE*, cf. supra) (l'homonymie avec la famille de *NODUS* "nœud" qui appartient à d'autres *frames* n'y joue pas un rôle déterminant)
- = lat. parlé tard. *avena* "rênes" (< *HABENAE*) ~ *avena* "avoine" (< *AVENA*)
- = l'exemple classique pour ce type: gasc. *gat* "coq" (< *GALLUS*) ~ *gat* "chat" (< *CATTUS*) → bigey "coq" (< *VICARIUS*) voire *hasâ* "id." (< *PHASIANUS*)
- un isolement morpho-sémantique: *SUPERBUS* "hautain" (survit comme emprunt savant "superbe" en français, et dans l'it. *superbo*) → fr. *hautain*, esp. *altivo* qui se rattachent à la famille très fréquente de *haut*, *alto*
- un (relatif) isolement morphologique: afr. *clore* → fr.mod. *fermer* (un verbe du 3^e groupe est remplacé dans la plupart de ses occurrences par un verbe du 1^{er} groupe).

Très souvent un mot ne disparaît pas intégralement d'une langue mais change de fréquence et de contexte d'utilisation, en passant d'une variété de la langue à une autre, sociolectale ou régionale. Les transformations onomasiologiques chevauchent alors celles qui parcourent l'espace géolinguistique.

6. Aspects diasystématiques et évaluatifs

6.1. La variation diatopique

La variation diasystématique concerne toutes les langues et tous les domaines linguistiques; elle est néanmoins particulièrement développée dans des langues historiques élaborées et son action implique le lexique bien plus que la grammaire qui garantit en revanche la cohésion d'un diasystème. Dans une société moderne, seule une partie réduite du vocabulaire est partagée par l'essentiel des locuteurs; la majorité des mots sont liés à des contextes d'usage ou à des connaissances particulières. Cette constatation s'accroîtrait encore en considérant le vocabulaire dialectal; mais puisque le français a connu, comme d'autres langues occidentales, une

rupture réelle entre les dialectes parlés et la langue écrite régionalisée depuis le MA (→ art. 214), il est sans doute plus judicieux d'exclure la variation dialectale de ce contexte.

La variation diatopique régionale est moins étudiée – et éventuellement même moins développée – en français que la variation dialectale mais elle est présente depuis les débuts de la langue écrite et constitutive pour les *scriptae* médiévales (Gleßgen / Thibault 2005, 12s.); la diffusion du français comme langue parlée dans la société aux époques moderne et contemporaine a accentué le développement d'éléments régionaux, notamment dans le lexique (→ art. 218). Cette variation est particulièrement riche dans la Francophonie européenne et extra-européenne (→ art. 238; Poirier 2005); elle intervient alors dans la constitution de variétés standard nationales (cf. pour la Suisse et le Québec le DSR et le DHFQ). Le français hors de France connaît de nombreux régionalismes de fréquence (cf. Thibault 2007) et de fortes variations dans les liens syntagmatiques existant entre les mots (Gleßgen 1997) ou dans la phraséologie, souvent à fins ludiques. La variation régionale concerne cependant tout autant le français en France, notamment par les nombreux régionalismes lexicaux (cf. le DRF de Pierre Rézeau) dont beaucoup sont aussi des régionalismes de fréquence. Les mots régionaux se répartissent entre des dénominations de concepts généraux (ex. "la serpillière", Lebsanft 2005, 293s., et, pour le même ex., → art. 164, 4.2.; cf. aussi: *cornet*, *sachet*, *poche*) et des concepts liés à une région donnée (comme *kouglof*; cf. comme modèle pour une région le DRFA).

Quand les régionalismes sont reconnus comme tels par les locuteurs – ce qui n'est pas toujours le cas – ils peuvent véhiculer une charge identitaire notable. Les régionalismes interviennent depuis le MA dans les échanges variationnels à travers le diasystème français: ils peuvent être empruntés d'une région à une autre, se dérégionaliser et entrer dans la langue générale ou encore servir de chaînon intermédiaire entre des langues alloglottes régionales (occitan, alsacien etc.) et le français général ou entre les dialectes galloromans et de français; les épïcêtres de diffusion des régionalismes sont surtout les grands centres urbains (cf. Rézeau 2007; → art. 218, 1.).

Quand un régionalisme dépasse d'une manière ou d'une autre son cadre d'usage

habituel, sa charge connotative augmente (par ex. *septante* est diatopiquement connoté en France mais non pas en Belgique, en Suisse ou au Canada); quand il est dérégionalisé, cette charge peut se fixer dans le contenu sémémique des mots en question (un *kouglof* est "une forme de brioche, produite en Alsace ou d'après des modèles alsaciens"), en cohérence avec le modèle sémantique des trois niveaux d'Andreas Blank (cf. supra 4.1.1.).

La recherche sur les régionalismes lexicaux du français s'est intensifiée depuis la dernière décennie du XX^e s., à commencer par la description de la variation actuelle; mais elle connaît encore de nombreux *desiderata*, notamment dans leur évolution historique.

6.2. La variation diastratique

La variation diastratique est présente, elle aussi, depuis les débuts de l'écrit français, ne serait-ce que par la présence massive de latinismes dans les textes; ces formes ressemblent en dernière instance à des emprunts variationnels, puisque le latin (la variété de haut prestige) était perçu par les scribes comme une forme de langue appartenant au même diasystème que le français (la variété de faible prestige).

Parallèlement à la diversification et la complexification de la société occidentale, les variétés sociolectales du français se sont développées, notamment depuis l'époque moderne. Le français familier doit sans doute sa genèse à un double mouvement de (1) dédialectalisation dans les grands centres urbains comme Paris et Lyon et de (2) diversification à partir des formes du français (pré-)standard. Les oppositions grammaticales et phonétiques sont presque aussi marquées entre le français courant et le français familier que les oppositions lexicales, ce qui s'explique sans doute par l'ancienneté de cette opposition et par son importance fonctionnelle (le français familier est largement complémentaire du français courant selon les contextes d'usage et fortement lié à l'oral). En revanche, les autres variétés diastratiques, telles que le français populaire, les formes de jargon ou encore le français soigné et les langages spécialisés se caractérisent surtout par des inventaires lexicaux très développés par rapport au français familier (les premiers) ou standard (les seconds).

La diversification diastratique du vocabulaire est fortement marquée en français

actuel (cf. l'ex. des formes *voiture*, *auto*, *automobile*, *bagnole*, *caisse*, *chiotte*, donné par Müller 1985, 227), même si la classification des différents niveaux de prestige linguistique n'est pas satisfaisante ni dans la lexicographie ni dans bon nombre d'études spécialisées. Par conséquent, les répertoires actuellement disponibles fournissent de nombreux exemples de termes familiers et populaires (→ art. 203; 206; 209); mais ils ne sont pas fiables du point de vue de la théorie variationnelle. La description historique des formes reste, elle-aussi, à développer, même si la plupart des termes populaires actuels ne remontent pas au delà de la fin du XIX^e s., début de la grande diversification sociale.

Un phénomène fréquent mais souvent mal cerné dans les études lexicologiques est l'emprunt variationnel, dans des usages autant occasionnels que lexicalisés; comme pour les formes régionales, un mot diastratiquement marqué et utilisé en dehors de son contexte habituel gagne une charge sémantique connotative supplémentaire (par ex. le mot *moufflet* utilisé dans un contexte de langue courante provoque un effet de style).

Dernière remarque: la variation diastratique s'exprime en français contemporain dans une large mesure à travers un vocabulaire qui est en même temps diaphasiquement marqué (→ art. 203).

6.3. La variation diaphasique

La variation diaphasique concerne essentiellement les modèles textuels et le lexique, parfois aussi des préférences grammaticales (→ art. 170, 5.; 197; 198). Son intensité dépend de la diversification des domaines du savoir et des relations sociales; sa croissance depuis le MA s'est donc largement intensifiée depuis la fin du XIX^e s. Dans les époques anciennes, elle fait surface, essentiellement, dans des termes liés à un savoir spécialisé et à certains genres textuels. Il existe de nombreuses études sur le vocabulaire du droit, de la médecine, de la religion, de l'économie, de l'agriculture ou des techniques mais aussi des différents genres littéraires au MA (→ art. 171, 5.; 175; 181; 185; 187); ces travaux renseignent sur des termes liés à certains contextes, même s'il manque souvent une théorisation de ce qui peut être perçu comme un marquage diaphasique dans les époques anciennes: dans les cas concrets, il faut toujours essayer de distinguer un regroupement purement sémantique (par ex., un mot qui appartient au monde de l'agriculture

mais qui est généralement connu et courant) et des termes plus spécialisés qui ont une charge communicative particulière (cf. Gleßgen 2006).

Depuis la formation des langages techniques et scientifiques, à partir du XVIII^e s., et l'intensification des terminologies, à partir de la fin du XIX^e s., la variation diaphasique englobe des ensembles quantitativement impressionnants de mots ou termes spécialisés, souvent à diffusion internationale (→ art. 70, 7./8.); les terminologies des différentes sciences dépassent de loin la quantité de mots naturels, liés à un savoir partagé (→ art. 192; 194); leur dimension numérique se place dans la sphère du million et non pas dans celle d'une centaine de milliers d'entrées.

Par ailleurs, la diversification de la société a produit, surtout au XX^e s., des ensembles lexicaux importants dans d'autres domaines comme ceux de la publicité, du sport ou, généralement, de la presse et des médias audio-visuels (→ art. 188; 189; 195; 196).

Le langage des jeunes et des banlieues, développé depuis les années 1960 et fortement intensifié depuis les années 1980 (→ art. 206; cf. supra 3.4.; 4.2.4.), représente une deuxième catégorie importante dans le vocabulaire diaphasiquement marqué. Si les langages techniques s'articulent sur la base du français courant ou standard, les langages périphériques reposent sur la base lexicale et grammaticale du français familier et populaire. La créativité lexicale des différentes formes de la périphérie sociale est extrêmement forte; elle mène à des emprunts à l'anglais, à l'arabe et aux formes argotiques plus anciennes ainsi qu'à des changements sémantiques et dérivationnels en grand nombre. Les mécanismes préférentiels de la formation de mots sont notamment l'abréviation ou la conversion (cf. supra 4.2.2.). Le vocabulaire généré ici – et élaboré dans des romans ou dans le rap – se diffuse fortement dans le français familier des dernières décennies, par des formations autant occasionnelles que lexicalisées (cf. les nombreux exemples → art. 206).

6.4. Aspects évolutifs et évaluatifs du diasystème

La transformation du vocabulaire à travers le temps crée une dimension chronologique à l'intérieur du diasystème. Les mots qui sont en train de se diffuser dans l'usage (les néologismes) tout autant que ceux qui sont en voie de disparition (les archaïsmes),

provoquent facilement des réactions émotionnelles ('beau-laid', 'bien-mauvais', cf. supra 3.5.); celles-ci disparaissent lors de l'intégration complète des néologismes (par ex., un anglicisme ancien comme *film* ou une abréviation comme *radar* 1943 ne provoquent pas de phénomènes évaluatifs). Dans ce sens, il existe un marquage diachronique du vocabulaire qui se greffe sur les trois catégories présentées jusqu'ici. Si un texte comporte beaucoup de néologismes ou d'archaïsmes, cela lui attribue une forte charge connotative comme lorsqu'il comporte beaucoup de régionalismes.

En résumé, les différentes formes de variation diasystématique sont constitutives pour de nombreux processus évolutifs dans la langue (cf. Gleßgen 2005); elles permettent de gérer de manière souple l'innovation, la diffusion et l'abandon d'éléments lexicaux. Par ailleurs, elles possèdent toutes un fort impact communicatif par leur marquage connotatif. La linguistique 'populaire' (*Laienlinguistik*) se concentre par conséquent souvent sur des mots marqués dans le diasystème, y inclus les emprunts récents et d'autres néologismes (→ art. 132, 6.1.; Schweickard 2005). De même, l'élaboration linguistique intervient toujours dans des segments définis du diasystème dont les formes nouvellement instaurées peuvent se diffuser ensuite dans la langue générale. La relation entre les mots diasystématiquement neutres et marqués n'a jamais été étudiée de façon systématique, même si le français fournirait, une nouvelle fois, une base d'observation exemplaire pour de telles études.

7. Perspectives

La lexicologie doit tenir compte de deux phénomènes contradictoires mais complémentaires: d'un côté, le vocabulaire d'une langue fonctionne en synchronie de façon articulée et organisée, d'un autre, chaque mot connaît une distribution très individuelle dans l'ensemble des variétés et contextes d'usage de la langue. Les perspectives de la recherche historique s'inscrivent dans ces deux tendances générales. Pour la première, il est nécessaire de gagner des vues d'ensemble sur les tendances évolutives du vocabulaire, distinguer les mots neutres de la langue générale des mots portant des marques diasystématiques aux différents moments historiques et identifier les grands mouvements de transformation dans le lexique; pour la seconde, chaque mot

doit être étudié en détail pour bien déterminer son contexte de genèse et d'usage, ses implications syntagmatiques, son évolution sémantique et sa diffusion dans le diasystème.

Les perspectives de la linguistique de corpus avec ses possibilités d'évaluation quantitative et statistique permettent pour la première fois d'évaluer le vocabulaire de la langue dans son ensemble; de telles études doivent néanmoins intégrer à tout moment les acquis et les méthodes de la lexicologie traditionnelle: une forme lexicale sans ancrage dans le diasystème, sans datation ou sans définition sémantique et contextuelle reste sans utilité pour l'histoire linguistique (cf. Gleßgen / Gouvert 2007).

La lexicologie historique peut réclamer une place à part entière à côté de la grammaire et de la phonétique historiques. La théorie variationnelle et la quantification linguistique, de même que la linguistique cognitive fournissent des apports nouveaux qui, alliés aux grandes traditions philologique et étymologique en romanistique permettent de développer la place de cette discipline trop longtemps considérée comme périphérique par les grammairiens.

8. Bibliographie

- Arveiller, Raymond, *Addenda au FEW XIX (Orientalia)*, ed. Max Pfister, Tübingen, 1999.
- Baldinger, Kurt (ed.), *Etymologien. Untersuchungen zu FEW 21–23*, 3 vol., Tübingen, 1988/2002.
- Blank, Andreas, *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*, Tübingen, 1997.
- , *Der 'Kopf' in der Romania und anderswo, ein metaphorisches (und metonymisches) Expansions- und Attraktionszentrum*, in: Gil, Alberto / Schmitt, Christian (eds.), *Kognitive und kommunikative Dimensionen der Metaphorik in den romanischen Sprachen*, Bonn, 1998, 11–32.
- , *Einführung in die lexikalische Semantik für Romanisten*, Tübingen, 2001.
- Brunet, Etienne, *Le vocabulaire français de 1789 à nos jours d'après les données du Trésor de la langue française*, 3 vol., Genève, 1981.
- , *Rabelais et son temps. Base de données interactive sur l'œuvre complète de Rabelais illustrée de quelques documents iconographiques*, 2004. <<http://ancilla.unice.fr/abelais.html>> / <<http://134.59.31.3/rabelais.html>> / <<http://www.uninec.fr/bcl>> (= Bases Corpus Langage) (17/03/2007).
- Buchi, Eva, *Les Structures du 'Französisches Etymologisches Wörterbuch'. Recherches métalexicographiques et métalexicologiques*, Tübingen, 1996.
- Chambon, Jean-Pierre / Carles, Hélène, *À propos du traitement des emprunts à l'occitan dans le TLF*, in: Rézeau 2007, 313–325.
- Enckell, Pierre / Rézeau, Pierre, *Dictionnaire des onomatopées*, Paris, 2003.
- Ernst, Gerhard, *Das Frequenzwörterbuch – ein Hilfsmittel der vergleichenden Sprachcharakteristik?*, in: id. / Stefenelli, Arnulf, *Sprache und Mensch in der Romania. Heinrich Kuen zum 80. Geburtstag*, Wiesbaden, 1979, 21–43.
- , *Konvergenz in Wortbildung und Semantik romanischer Sprachen am Beispiel der Nomina qualitatis*, in: RK VIII (1995), 65–84.
- Geeraerts, Dirk, *Theories of lexical semantics: a cognitive perspective*, Oxford, 2004.
- Gleßgen, Martin-D., *Spielarten des Sprachwandels: Französische Leserbriefe aus Marokko als sprachwissenschaftliche Quelle*, ZfSL 107 (1997), 2–35.
- , *Diskurstraditionen zwischen pragmatischen Regeln und sprachlichen Varietäten*, in: Schrott, Angela / Völker, Harald (eds.), *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*, Göttingen, 2005, 207–228.
- , *Vergleichende oder einzelsprachliche historische Textwissenschaft*, in: RK XX (2006), 319–340.
- , *Linguistique romane. Domaines et méthodes en linguistique française et romane*, Paris, 2007.
- Gleßgen, Martin-D. / Gouvert, Xavier, *La base textuelle du 'Nouveau Corpus d'Amsterdam': ancrage diasystématique et évaluation philologique*, in: Kunstmann, Pierre / Stein, Achim, *Le Nouveau Corpus d'Amsterdam. Actes de l'atelier de Lauterbad (23–26 février, 2006)*, Stuttgart, 2007, 51–84.
- Gleßgen, Martin-D. / Thibault, André (eds.), *La lexicographie différentielle du français et le 'Dictionnaire des régionalismes de France'. Actes du Colloque en l'honneur de Pierre Rézeau pour son soixante-cinquième anniversaire*, Strasbourg, 2005.
- Gougenheim, Georges, et al., *L'élaboration du français élémentaire*, Paris, 1956.
- , *L'élaboration du français fondamental (1^{er} degré)*, Paris, 1964.
- Grevisse, Maurice, *La bon usage*, Paris / Louvain-la-Neuve, ¹³1993 (1936; réimpr. 2001).
- Höfler, Manfred, *Dictionnaire des anglicismes*, Paris, 1982.
- Juilland, Alphonse / Brodin, Dorothy / Davidovitch, Catherine, *Frequency dictionary of French words*, Den Haag, 1970.
- Kleiber, Georges, *La sémantique du prototype. Catégories et sens lexical*, Paris, 1990.
- Koch, Peter, *Lexical typology from a cognitive and linguistic point of view*, in: HSK 20/2 (2001), 1142–1178.
- Laca, Brenda, *Die Wortbildung als Grammatik des Wortschatzes*, Tübingen, 1986.

- Lebsanft, Franz, *Studien zu einer Linguistik des Großes*, Tübingen, 1987.
- , *Régionalismes et 'culture de la langue' dans le monde francophone*, in: Gleßgen / Thibault 2005, 289–297.
- Lebsanft, Franz / Gleßgen, Martin-D. (eds.), *Historische Semantik in den romanischen Sprachen*, Tübingen, 2004.
- Lüdtke, Jens, *Gemeinromanische Tendenzen IV. Wortbildungslehre*, in: LRL 2/1 (1996), 235–272.
- Müller, Bodo, *Le français d'aujourd'hui*, Strasbourg, 1985 (ed. orig. all. *Das Französische der Gegenwart*, Heidelberg, 1975).
- Pfister, Max, *Der 'Kopf' im Italienischen und in den romanischen Sprachen*, in: Lebsanft / Gleßgen 2004, 141–151.
- Pfister, Max / Lupis, Antonio, *Introduzione all'etimologia romanza*, Catanzaro, 2001.
- Poirier, Claude, *La dynamique du français à travers l'espace francophone à la lumière de la base de données lexicographiques panfrancophone*, RLiR 69 (2005), 483–516.
- Pusch, Claus / Kabatek, Johannes / Raible, Wolfgang (eds.), *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft*, Tübingen, 2005.
- Quadri, Bruno, *Aufgaben und Methoden der onomasiologischen Forschung*, Bern, 1952.
- Raible, Wolfgang, *Zur Einleitung*, in: Stimm, Helmut / Raible, Wolfgang (eds.), *Zur Semantik des Französischen*, Wiesbaden, 1983, 1–24.
- Rézeau, Pierre (ed.), *Richesses du français et géographie linguistique. Recherches lexicographiques sur les variétés du français de France*, Bruxelles, 2007.
- Schmitt, Christian, *Wörter und Sachen*, in: LRL 1/1 (2001), 235–292.
- Schweickard, Wolfgang, *'Il se croit de la côte d'Adam'. Reflexe der biblischen Vorstellungswelt im übertragenen Sprachgebrauch der romanischen Sprachen*, in: RK III (1990), 23–45.
- , *'Deonomastik'. Ableitungen auf der Basis von Eigennamen im Französischen unter vergleichender Berücksichtigung des Italienischen, Rumänischen und Spanischen*, Tübingen, 1992.
- , *Englisch und Romanisch*, in: LRL 7 (1998), 291–310.
- , *Glanz und Elend der Sprachpflege. Der Umgang mit Anglizismen in Frankreich, Italien und Deutschland*, in: RK XVIII (2005), 177–191.
- Stefenelli, Arnulf, *Geschichte des französischen Kernwortschatzes*, Berlin, 1981.
- , *Das Schicksal des lateinischen Wortschatzes in den romanischen Sprachen*, Passau, 1992.
- , *Gesamtromanische Tendenzen VIII. Lexikon und Semantik*, in: LRL 2/1 (1996), 368–386.
- Thibault, André, *Banques de données textuelles, régionalismes de fréquence et régionalismes négatifs*, in: ACILPR XXIV (section 2), 2007/1, 467–480.
- Wolf, Heinz Jürgen, *Französische Sprachgeschichte*, Heidelberg / Wiesbaden, ²1991.
- Wunderli, Peter, *Französisch: Lexikologie und Semantik*, in: LRL 5/1 (1990), 94–112.

Martin-Dietrich Gleßgen, Zürich